Universidade de Aveiro
2012
Departamento de Electrónica, Telecomunicações e Informática

Nelson Miguel Martins Coelho

Profiling de Tráfego Inter-Operador Baseado em Análise Multi-Escalar

Universidade de Aveiro
2012
Departamento de Electrónica, Telecomunicações e Informática

Nelson Miguel Martins Coelho

Profiling de Tráfego Inter-Operador Baseado em Análise Multi-Escalar

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Mestrado Integrado em Engenharia Eletrónica e Telecomunicações, realizada sob a orientação científica do Doutor Paulo Jorge Salvador Serra Ferreira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e sob a co-orientação científica do Doutor António Manuel Duarte Nogueira, Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro da Universidade de Aveiro

Dedicatória

Dedico este trabalho aos meus pais e à minha irmã, por todo o apoio, conselhos e incentivo que deram ao longo do meu percurso académico.

o júri

presidente	Professor Doutor Amaro Fernandes de Sousa Professor auxiliar da Universidade de Aveiro
vogal	Professor Doutor Joel José Puga Rodrigues Professor auxiliar da Universidade da Beira Interior
orientador	Professor Doutor Paulo Jorge Salvador Serra Ferreira Professor auxiliar da Universidade de Aveiro
co-orientador	Professor Doutor António Manuel Duarte Nogueira Professor auxiliar da Universidade de Aveiro

Agradecimentos

Quero agradecer aos meus pais e à minha irmã, por todo o apoio e encorajamento ao longo da minha vida académica.

Agradeço aos meus orientadores por todo o apoio e orientação que prestaram ao longo da elaboração deste trabalho, ajudando a ultrapassar as dificuldades e a encontrar a direção certa.

A todos os meus amigos, por todo o apoio, acompanhamento e bons momentos proporcionados, o meu muito obrigado.

Agradeço também aos meus colegas de laboratório, pela ajuda prestada e pelo espírito de camaradagem sempre presente durante os dias de trabalho.

Palavras-chave

Perfil de tráfego, fluxos, *wavelets*, escalogramas,diferenciação multi-escalar, perfil de utilizador, qualidade-de-serviço,aplicações.

Resumo

O acesso à Internet nos últimos anos generalizou-se de tal forma que tornou-se um bem essencial no nosso dia-a-dia, seja para trabalho ou lazer. Contudo, esta rede global acarreta uma grande complexidade ao nível da gestão e monitorização de todas as suas ligações, tornando-se necessário assegurar qualidade-de-serviço (QdS) em toda a rede, de modo a garantir uma utilização eficaz por parte dos utilizadores particulares e empresariais. O aparecimento constante de novas aplicações introduziu a necessidade da implementação de protocolos de comunicação com diferentes requisitos, de acordo com a aplicação a que se destinam. Daí a necessidade de se mapear o tráfego na rede, associando-o à respetiva aplicação, de modo a se poder efetuar o *profiling* dos utilizadores da Internet.

Esta dissertação surge motivada pela necessidade de identificar e caracterizar fluxos de dados ao nível da rede core, de modo que o seu mapeamento permita identificar os requisitos destes fluxos, otimizando a pré-alocação de recursos de rede. As dinâmicas de utilização de rede são identificadas recorrendo a análise multi-escalar do tráfego.

Os objetivos desta dissertação consistem em identificar diferentes perfis de tráfego inter-operador, recorrendo a análise multi-escalar do tráfego e mostrando que é possível diferenciar fluxos de dados na rede core, associando-os a diferentes aplicações e identificando requisitos de recursos. Sabendo de antemão as necessidades de recursos das várias aplicações é possível ajustar os parâmetros de QdS das mesmas, permitindo aos operadores otimizar o desempenho da rede e do serviço.

.

Keywords Traffic profiling, flow,wavelets, scalograms, multiscale analysis, user profiling, quality-of-service, applications.

Abstract

The access to Internet has widespread in the last few years in such a thriving way that it became something essential in our daily routine, both at work and leisure. However, the fact that it is a global network implies great complexity to administrate and control every single connection. Therefore, it's necessary to ensure quality-of-service (QoS) across the entire network, in order to guarantee an effective utilization from individual and business users. The continued appearance of new applications introduced the need for the implementation of communication protocols with different requirements, according with the respective application. Hence the need to map the traffic in the network, associating it with the respective application, in order to perform the profiling of Internet users.

This dissertation is motivated by the need to identify and characterize data flows in the core network, so that their mapping can help identify requirements of these flows, optimizing the pre-allocation of network resources. The network utilization dynamics are identified resorting to multiscale traffic analysis.

The objectives of this dissertation consist in identifying different traffic profiles with multiscale traffic analysis and showing that is possible to differentiate data flows in the core network, associating them to different applications and identifying requirements of resources. Knowing beforehand the resource requirements of various applications, it's possible to improve their QoS parameters, allowing the providers to optimize the performance of both network and service.

Índice

Lista de Figuras

Lista de Tabelas

Acrónimos e Siglas

AV – Audio-Vídeo

Bps – Bytes por Segundo

CWT – Continuous Wavelet Transform

DiffServ – Differentiated Services

DWT – Discrete Wavelet Transform

FTP – File Transfer Protocol

FT – Fourier Transform

FTTH – Fiber To The Home

HTTP – Hypertext Transfer Protocol

HTTPS - Hypertext Transfer Protocol Secure

IMAP – Internet Message Access Protocol

IntServ – Integrated Services

ISP – Internet Service Provider

IPTV – Internet Protocol Television

KBps - Kilobyte por Segundo

LSP – Label Switched Path

MPLS – Multi-Protocol Label Switching

MSNP – Mobile Status Notification Protocol

POP3 – Post Office Protocol 3

QdS – Qualidade de Serviço

RSVP – Resource Reservation Protocol

SMTP – Simple Mail Transfer Protocol

SP – Service Provider

SSH – Secure Shell

TCP – Transmission Control Protocol

UDP – User Datagram Protocol

UTC – Coordinated Universal Time

VoIP – Voice over Internet Protocol

1 Introdução

Desde a sua abertura à utilização do cidadão comum, a Internet experimentou uma evolução meteórica, tanto ao nível da quantidade de utilizadores como ao nível das ferramentas e serviços disponibilizados online que poderiam ser utilizados. Se inicialmente as pessoas acediam à Internet para consultar e-mails de texto ou visualizar páginas Web básicas desenvolvidas em HTML, a evolução da tecnologia (computadores mais potentes, ligações de banda larga cada vez mais rápidas, maior capacidade dos servidores) permitiu que fosse possível aceder a cada vez mais e melhores serviços e aplicações online.

A possibilidade de ver canais TV, comunicar com pessoas (através de chat ou VoIP), fazer compras ou partilhar fotos e vídeos através das redes sociais atraíram cada vez mais pessoas para a rede; em 1995, 16 milhões de pessoas em todo o mundo acediam à Internet enquanto em Março de 2012 esse número ascendia aos 2280 milhões (cerca de 32.7% da população mundial [1]). Para satisfazer esta procura imensa, existem diversos operadores (*Internet Service Providers*, ISPs) por país responsáveis por garantir aos utilizadores acesso à Internet e que estando ligados entre si fazem da Internet um espaço realmente global.

O surgimento em força do conceito Web 2.0 revolucionou a forma como as pessoas acedem à Internet assim como a disposição dos conteúdos, pois as pessoas passaram a ter uma participação cada vez mais ativa na produção dos conteúdos e da informação. Este conceito foi importante para a aparição das redes sociais e do *cloud computing*, serviços com relevância crescente na rede global. Esta revolução de conteúdos, aplicações e serviços impôs o desenvolvimento de novos protocolos, devido aos diferentes requerimentos associados a cada aplicação.

Assim, o crescimento da complexidade na Internet trouxe desafios ao nível da gestão e da compreensão destas novas aplicações e do tráfego por elas gerado. Consequentemente, os ISPs depararam-se com padrões de tráfego desconhecidos e novos requisitos de recursos. Tendo em conta a necessidade de melhorar constantemente o serviço prestado aos seus clientes, é extremamente importante conseguir efetuar um mapeamento rigoroso destas novos padrões de tráfego e associá-los à respetiva aplicação para possibilitar a elaboração de perfis de tráfego. Estes perfis permitem identificar os recursos necessários para uma aplicação específica, tal como a alocação de largura de banda ou os requisitos de tempo de espera e atraso. Esta divisão do tráfego em diferentes classes de serviço com características específicas permite uma melhor distribuição dos recursos necessários para a execução de cada aplicação, facilitando uma melhor Qualidade-de-Serviço (QdS).

Os gestores de rede também têm muito a ganhar com o mapeamento do tráfego, pois permite-lhes melhorar os índices de performance da rede, segurança, diferenciação do serviço e a gestão dos recursos da rede. Sabendo de antemão os perfis dos utilizadores que acedem à rede, é possível prever o seu comportamento futuro e assim alocar previamente recursos da rede de modo a que a performance desta seja a melhor possível.

Várias metodologias surgiram ao longo dos anos com diferentes abordagens à questão da caracterização de perfil (*profiling)* de tráfego, baseadas sobretudo na análise estatística do tráfego e na investigação aprofundada do conteúdo dos pacotes. Contudo, o surgimento de aplicações de maior complexidade assim como as preocupações com a encriptação de tráfego e proteção da privacidade dos dados dos utilizadores obrigou essas metodologias a acompanharem a evolução da própria Internet. Portanto, são necessárias novas abordagens à análise do tráfego na Internet que consigam tornear

estes entraves à sua performance. Esta dissertação aborda essa necessidade, comparando diferentes metodologias e tentando perceber as suas mais-valias e os itens que requerem melhorias.

1.1 Motivação

Uma das motivações deste trabalho prende-se com o facto da crescente complexidade de aplicações e serviços na Internet requerer uma abordagem que caracterize o tráfego na Internet tendo em conta os padrões dos fluxos de dados. O aparecimento de novas aplicações multimédia cada vez mais complexas obrigaram os ISPs a perceber que recursos QdS são mais relevantes para o melhoramento do serviço prestado ao nível destas aplicações. As aplicações relacionadas com redes sociais e a partilha de ficheiros (P2P) têm tido um grande crescimento nos últimos anos e carecem de estudo, dada a sua constante evolução e transformação.

Pretende-se identificar e caracterizar fluxos de dados ao nível da rede core, de modo que o seu mapeamento permita identificar os requisitos destes fluxos de tráfego. Assim, conhecendo o comportamento do tráfego gerado por diferentes aplicações, será possível suprir as suas necessidades a curto prazo, com uma distribuição de recursos mais personalizada e eficaz, proporcionando assim uma melhor QdS nessas aplicações e libertando recursos que sejam necessários para diferentes aplicações e funcionalidades da rede.

O tráfego gerado por uma aplicação na Internet tem características muito próprias, pois é delineado por eventos e procedimentos que ocorrem em diferentes componentes de frequência do espetro de frequência. Estes componentes envolvem eventos que ocorrem ao longo de toda a gama de frequência: frequências baixas, intermédias e altas. Efetuando uma análise multi-escalar, investiga-se as dinâmicas do tráfego na rede e avaliando os diferentes componentes de frequência é possível descobrir que tipo de eventos estão relacionados com um certo tipo de tráfego. Assim, otimiza-se a pré-alocação de recursos de rede tendo em conta os eventos e mecanismos associados a determinado fluxo de dados. Tendo em conta os diferentes componentes de frequência existentes nos fluxos de dados, poderão definir-se regiões ao longo do espetro de frequência e mapeando essas mesmas regiões os fluxos de dados poderão ser associados a diferentes aplicações.

As metodologias de análise de tráfego existentes normalmente analisam o fluxo completo para obter estatísticas sobre o tráfego em questão ou então verificam o conteúdo dos pacotes do dito tráfego, o que vai contra questões de ordem legal e políticas restritivas adotadas pelos ISPs e não é eficiente para analisar tráfego encriptado. Assim, o paradigma proposto para a caracterização de perfis de tráfego pretende efetuar monitorização de baixo nível e análise assente no tratamento estatístico da camada 3, ou seja, analisando sobretudo o cabeçalho IP dos pacotes dos fluxos capturados, de modo a evitar os problemas relacionados com a privacidade e a encriptação dos dados.

1.2 Objetivos

O objetivo principal do trabalho a executar no âmbito desta dissertação é efetuar *profiling* de tráfego recorrendo à análise multi-escalar dos fluxos de dados capturados e tentar perceber se a multiplexagem do tráfego na orla da rede se mantém na rede core. A diferenciação do tráfego analisado e a sua caracterização permitem examinar o comportamento de cada classe de serviço em estudo. Assim, será possível perceber os parâmetros de QdS mais importantes para uma melhor execução de cada aplicação. Mas para atingir este fim são imperativos alguns passos intermédios.

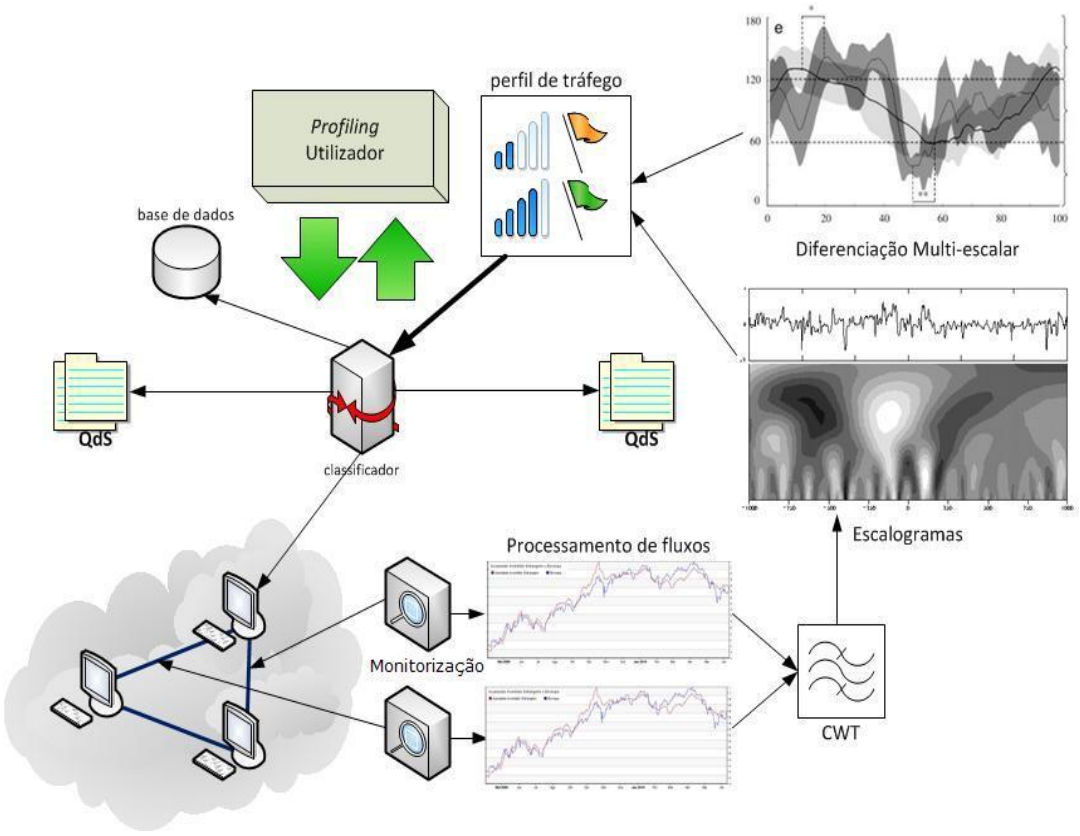

Figura 1.1 – Arquitetura de classificação de tráfego com suporte para QdS (baseada em [44]).

O primeiro objetivo a cumprir será diferenciar os fluxos de dados capturados na rede core através da sua caracterização e associá-los a diferentes aplicações, tendo em conta a análise multi-escalar destes fluxos. Depois de mapeados estes fluxos, será possível identificar os requisitos de recursos dos mesmos, tanto no presente como no futuro, de forma a melhorar a QdS das diferentes aplicações.

Outro dos objetivos será definir uma metodologia para a aplicação da caracterização de perfis de tráfego desenvolvida ao longo deste trabalho. A Figura 1.1representa de forma genérica a proposta para uma arquitetura de classificação de tráfego com suporte para QdS. O tráfego agregado é monitorizado em vários monitores, que extraem fluxos para serem divididos segundo os seus componentes de frequência e tempo. Dessa decomposição resultam escalogramas e é possível construir um perfil para o tráfego gerado por cada aplicação, com os seus diferentes componentes de frequência. O classificador associa tráfego capturado e desconhecido a diferentes classes de serviço e seus respetivos parâmetros QdS. Assim, recorrendo à base de dados das assinaturas

digitais, o classificador consegue associar cada novo fluxo de dados à respetiva classe de serviço e atualiza também os parâmetros de QdS. No final deste processo, esta informação é transferida para o módulo de *Profiling* de Utilizador que atualiza o perfil de cada utilizador de acordo com a informação que vai recebendo ao longo do tempo. Este procedimento será apresentado com mais detalhe nos capítulos 2, 3 e 4.

1.3 Estrutura

Esta dissertação está estruturada em seis capítulos:

- Capítulo 1: Neste capítulo é feita a introdução geral ao trabalho desenvolvido nesta dissertação; é explicada a motivação desta dissertação; são expostos os objetivos deste trabalho e é apresentada a estrutura geral da dissertação.

- Capítulo 2: Neste capítulo é feito o enquadramento geral deste trabalho. São explicados os conceitos e estratégias relacionadas com o *profiling.* São apresentados diferentes métodos de classificação de tráfego, apontando diferenças e semelhanças entre eles. É feita uma abordagem de alto nível sobre parâmetros e mecanismos QdS.

- Capítulo 3: Neste capítulo são apresentados os conceitos de *wavelet* e escalograma. Faz-se a distinção entre *Continuous Wavelet Transform* (CWT) e *Dicrete Wavelet Transform* (DWT) e discute-se a sua importância para a análise multi-escalar. Explica-se a diferenciação multi-escalar, nomeadamente no que diz respeito à análise do desvio padrão da energia dos fluxos e a divisão em várias escalas de frequência e tempo.

- Capítulo 4: Neste capítulo é feita a descrição das capturas de tráfego utilizadas neste trabalho, a forma como foram recolhidos os dados, são apresentados os protocolos estudados e são explicados os procedimentos para o processamento dos dados e posterior análise.

- Capítulo 5: Neste capítulo apresentam-se, para cada protocolo estudado, os resultados obtidos. Através da comparação dos resultados obtidos, procura-se tirar conclusões sobre os vários aspetos em análise para que sejam levadas em conta em trabalhos futuros nesta área. Faz-se uma reflexão sobre a forma como o trabalho correu, o que pode ser melhorado e o que pode ser feito em possíveis trabalhos futuros.

- Capítulo 6: Neste capítulo retiram-se conclusões sobre os resultados obtidos e faz-se uma reflexão sobre a forma como o trabalho decorreu, o que pode ser melhorado e o que pode ser feito em trabalhos futuros.

2 Enquadramento do Trabalho

A evolução das infraestruturas terrestres de suporte das redes de telecomunicações ajudou a potencializar os serviços disponibilizados pela Internet; seria impraticável tentar visualizar vídeos online utilizando uma ligação com velocidade de 56 KBps (Kbytes por segundo), por exemplo. Assim, toda esta evolução tanto na velocidade das ligações de acesso à Internet como nos terminais de acesso possibilitou que cada vez mais serviços passassem a estar disponíveis online, tornando a Internet algo banal e ao mesmo tempo indispensável na rotina das pessoas.

A evolução da tecnologia e das infraestruturas das redes de acesso ao longo dos anos modelaram a forma como acedemos à internet: dos modems de 56 KBps que utilizavam as linhas telefónicas às ligações de banda larga; das ligações analógicas à fibra ótica com linhas dedicadas para cada cliente (*Fiber To The Home*, FTTH) capazes de fornecer VoIP, CATV e tráfego IP entre outros serviços de rede. Assim, foi possível o surgimento e evolução de mais serviços e ferramentas acessíveis online, originando uma crescente concorrência entre os prestadores de serviços, desejosos de captar o maior número de utilizadores possível, dado o grande crescimento na adesão à Internet globalmente.

A possibilidade de realizar vídeo chamadas e comunicar com outras pessoas utilizando programas de *Instant Messaging* (MSN Messenger, Skype) ou *Voice over IP* (VoIP) influenciaram em grande medida o tempo que as pessoas passam a navegar na Internet. Contudo, o aumento gradual do tráfego, principalmente no que concerne ao proveniente de terminais móveis como os *tablets* e os *smartphones* (em 2011, 22% do tráfego total global foi proveniente de dispositivos que não computadores [2]), coloca problemas aos ISPs ao nível da sustentabilidade das ligações que oferecem aos seus clientes. Espera-se que o tráfego na Internet tenha uma taxa de crescimento anual a nível mundial de 29% entre 2011 e 2016. [2]

Os utilizadores esperam sempre as melhores condições de serviço por parte do seu ISP. As ofertas de tráfego ilimitado, o gradual aumento nas velocidades de acesso à Internet e a crescente necessidade de mais largura de banda - devido sobretudo ao crescimento do consumo de conteúdos de vídeo e da partilha e alojamento de ficheiros volumosos online - colocaram novos desafios aos ISPs para garantir as melhores condições aos utilizadores, tendo em conta que cada utilizador tem necessidades e interesses específicos. Para resolver esta questão nem sempre é possível aumentar a capacidade do serviço, pois isso poderá ser dispendioso; como tal, é importante perceber os padrões de atividade dos utilizadores e acompanhar o comportamento dos fluxos de tráfego de diferentes serviços, de modo a encontrar um compromisso para que os recursos da rede sejam melhor aproveitados e para que o serviço prestado aos clientes seja o mais eficaz possível.

2.1 Estratégias de *Profiling*

O crescimento gradual da Internet implica o aparecimento praticamente contínuo de novos serviços e aplicações, que requerem o desenvolvimento de novos protocolos de suporte. Cada protocolo tem os seus requisitos, o que leva ao surgimento na rede de diferentes padrões e parâmetros de funcionamento. Tendo em conta a crescente exigência dos clientes para ter a melhor performance possível ao aceder online, torna-se extremamente importante que os SP façam um mapeamento completo do tráfego e que

desenvolvam perfis de utilizador, de modo a perceber melhor os hábitos dos utilizadores. Estes perfis devem ter em conta características da rede como a performance, o consumo de recursos, segurança ou diferenciação de serviços. [3]

Analisando os perfis de utilizador, os SP terão mais capacidade para interpretar os requerimentos de cada classe de serviço ao nível da largura de banda, tempos de resposta e atraso, o que lhes permite garantir parâmetros de QdS mais eficientes. Uma melhor alocação dos recursos de rede irá contribuir para que estes não se esgotem de forma desnecessária, aumentando assim a capacidade de resposta e a performance da rede. Com os perfis de utilizador será possível prever de forma mais segura os requisitos de cada utilizador numa utilização futura, permitindo uma redistribuição atempada dos recursos e não prejudicando o rendimento geral da rede; assim, no caso de haver muitos pedidos diferentes simultaneamente, a rede estará melhor preparada para responder aos mesmos e a probabilidade de haver saturação ou esgotamento da rede diminui.

Tendo em conta a natureza do tráfego IP, efetuar o mapeamento do tráfego e posterior caracterização do perfil do mesmo é uma tarefa altamente complexa, não sendo alheia a constante evolução tecnológica que verificamos atualmente. Ao longo dos anos, têm sido propostos vários métodos de classificação do tráfego, utilizando abordagens variadas, mas a crescente complexidade das novas aplicações integradas na Internet obrigaram que estes métodos sofressem algumas modificações de modo a acompanhar essa evolução. Consequentemente, convém fazer a distinção entre a caracterização de perfil de utilizador (*user profiling)* e caracterização de perfil de tráfego (*traffic profiling)*.

Traffic Profiling serve como base da gestão e planeamento da capacidade de resposta da rede. Geralmente, para executar estas ações de *profiling* os dispositivos de rastreio e monitorização de tráfego são colocados nas orlas da rede. Estes dispositivos operam em modo passivo, pois apenas analisam o tráfego que circula por eles, sem o alterar ou injetar tráfego alternativo. Desta forma, é possível obter estatísticas sobre os fluxos como o tamanho dos pacotes, a taxa de perda de pacotes, débito ou endereços IP de fonte e destino. Para efetuar a caracterização dos fluxos geralmente recorre-se a mecanismos estatísticos como a média ou o desvio padrão. A obtenção das estatísticas de vários fluxos implica o consumo intensivo de recursos do sistema, nomeadamente ao nível do desempenho dos dispositivos de monitorização de tráfego.

K. Ramantas et al. ([4]) introduziram o conceito de *traffic sampling*, que consiste em efetuar as análises de tráfego em subfluxos de pacotes, economizando no consumo de recursos. Neste caso, a escolha dos pacotes amostrados é perfeitamente aleatória. Os fluxos de amostragem são selecionados através de um processo aleatório de Bernoulli e todos os pacotes dos fluxos escolhidos são usados no processo de *profiling*. Os fluxos de amostragem são guardados em tabelas de dispersão (*hash tables*).

Garcia et al. ([5]) caracterizaram os pacotes de cada fluxo de tráfego analisado tendo em conta vários parâmetros: endereço IP de origem e destino, tamanho do pacote, tempo de chegada entre pacotes, carga de bytes no canal, carga de pacotes no canal e protocolos nos pacotes. Neste trabalho, os autores analisam o número de ligações entre os endereços de origem e destino calculando o total de pacotes e bytes que circulam entre cada par origem / destino. As estatísticas dos parâmetros anteriormente referidos serão analisadas de três diferentes perspetivas: usando a escala do tempo como uma linha do tempo contínua, cujo início coincide com o começo da amostragem dos fluxos; usando uma janela de tempo com comprimento de w segundos, pois os valores medidos alteram-se com o tempo e referem-se apenas a uma janela de w segundos; usando um pacote de amostragem a cada t segundos.

No estudo desenvolvido por K.C. Claffy et al. ([6]), o *profiling* do tráfego é feito tanto para fluxos individuais como para agregados de fluxos. Os autores optaram por este método pois as estatísticas obtidas sobre os agregados de fluxos influenciam o consumo de memória e outros recursos de suporte do sistema. A definição dos fluxos foi baseada na transmissão de pacotes em *endpoints* específicos na camada de rede. Esta metodologia pode ajudar a perceber como os parâmetros de fluxos afetam o design de arquiteturas de redes.

O design do sistema de *profiling* desenvolvido por K. Xu et al. ([7]) está assente em quatro características importantes em qualquer sistema de *profiling*: escabilidade (capacidade de processar vários fluxos em pouco tempo), robustez (capacidade de continuar a trabalhar mesmo se surgir tráfego anómalo), modularidade (cada módulo do sistema executa uma tarefa específica e comunica adequadamente com os outros módulos) e usabilidade (deve ser um sistema de fácil configuração). Neste sistema são construídos perfis de comportamento de *hosts* e aplicações de rede. Este método consiste em três passos: extração de clusters (fluxos com valores semelhantes numa determinada dimensão) relevantes, classificação automática de comportamento (agregação de tráfego com padrões comportamentais semelhantes em classes) e modelação estrutural (caracterização de diferentes dimensões num cluster). Este método de caracterização de perfil usa fluxos *5-tuple*, ou seja, endereço IP de origem (srcIP), endereço IP de destino (dstIP), porto de entrada (srcPrt), porto de saída (dstPrt) e protocolo.

O perfil de um utilizador pode ser descrito genericamente como o conjunto dos seus interesses, preferências e comportamentos[8]. Para a construção do perfil dos hábitos do utilizador é necessário compilar tráfego e dados que sejam identificáveis com o comportamento habitual de navegação do utilizador. Pode então dizer-se que *user profiling* corresponde ao processamento dos dados recolhidos sobre um utilizador e posterior interpretação dos mesmos de modo a que seja possível tentar prever o comportamento futuro do utilizador assim como os requisitos necessários para suportar a sua atividade. Todo este processo servirá para que a rede possa responder de forma mais eficaz aos requisitos de cada utilizador, o que permite que os parâmetros de QdS se aproximem do desejado. É impossível desenvolver um modelo de *user profiling* que tenha a descrição completa de todos os possíveis comportamentos de um utilizador, pois estes comportamentos variam com o tempo e existe a possibilidade de surgirem padrões novos e hábitos antigos simplesmente deixarem de existir. As definições de *user profiling* variam conforme as especificidades de cada método de classificação, assim como os resultados pretendidos. Portanto, no âmbito desta dissertação define-se perfil como um conjunto de aplicações, serviços e programas aos quais cada utilizador acede e interage e os requisitos necessários para suportar estas operações.

Iglesias et al. ([8]) apresentaram um modelo de *user profiling* que representa o comportamento de um utilizador como uma distribuição adaptada do seu comportamento, ou seja, dos eventos que o seu comportamento cria. A metodologia deste modelo tem a particularidade de atualizar os perfis dos utilizadores à medida que estes vão criando novos eventos. Este modelo tem várias ações sequenciais para efetuar o *clustering* automático, design do classificador e a classificação dos perfis dos utilizadores: primeiro são criados os perfis de utilizador, tendo em conta as sequências de comandos que estes digitam; o sistema de classificação é preparado para a eventualidade dos perfis dos utilizadores serem atualizados; finalmente, os perfis são classificados tendo em conta os protótipos de classes de serviço existentes na arquitetura deste modelo.

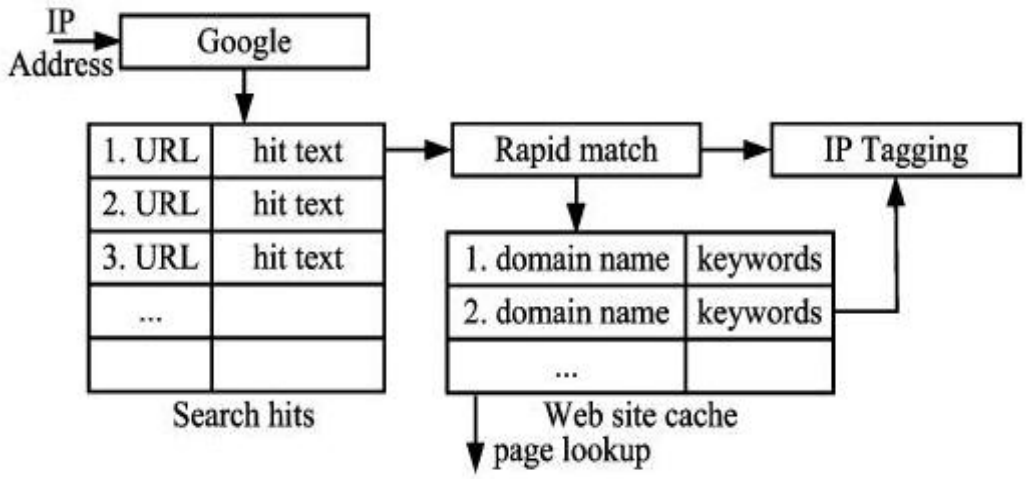

Figura 2.1 – Modelo de *Profiling Web-based.* Gera rótulos (*tags*) para os endereços IP com base em informação encontrada no Google. (Figura obtida de [9])

O modelo apresentado por I. Trestian et al. ([9]) caracteriza os hábitos de vários *endpoints*, combinando informação existente na Web sobre os mesmos. A Figura 2.1 esquematiza o modelo de caracterização de tráfego: um *rule generator* que opera sobre o motor de busca Google e um IP *tagger* que identifica *endpoints* com certas características. Este método de *profiling* envolve três passos: *Rule generation*, Classificação Web e *IP tagging*. O primeiro passo consiste em procurar certos endereços IP através do Google e agrupar os resultados com uma determinada classe de serviço. Existe uma tabela que associa classes de serviço a certas palavras-chave, de modo que quando essas palavras-chave surgem nos resultados da procura os endereços IP sejam facilmente encaminhados para a classe de serviço respetiva. O segundo passo consiste na classificação dos *endpoints* recorrendo à Web e às palavras-chave previamente referidas. O último passo consiste em etiquetar os endereços IP, recorrendo à informação previamente recolhida.

2.2 Métodos de Classificação de Tráfego

O crescimento e implementação da Internet como um veículo de transmissão de conteúdos de massas despoletou várias disputas a nível político, económico e legal. Alguns desses conflitos relacionam-se diretamente com a temática da classificação de tráfego: há a questão da partilha legal de ficheiros, que opõe a comunidade de partilha de ficheiros e os representantes dos detentores da propriedade intelectual (por exemplo, a *Recording Industry Association of America* [10] e a *Motion Picture Association of America* [11]); a eterna batalha dos hackers frente às companhias de software de segurança e a querela da neutralidade na rede entre os ISP e os fornecedores de serviços e aplicações [12]. Portanto, percebe-se as limitações de quem pretende classificar tráfego, pois muitas vezes o tráfego analisado pode incluir dados dos utilizadores. Assim, as técnicas de classificação estão consignadas a classificar tráfego recorrendo ao cabeçalho IP dos pacotes, tendo em conta estatísticas das camadas 2 e 3, respetivamente ligação de dados e rede) [13].

A capacidade de classificar aplicações com eficiência influencia positivamente várias tarefas ao nível da gestão e performance das redes como por exemplo a diferenciação de serviço, a definição dos parâmetros de QdS e a segurança. Um ISP que

consiga associar o tráfego de dados que circula na sua rede à sua respectiva aplicação de Internet original será capaz de agrupar tráfego com as mesmas características e optimizar a alocação de recursos conforme as necessidades de cada aplicação; também será mais fácil a deteção de ameaças à rede como *worms* ou *botnets.* [14]

Ao longo dos anos, surgiram vários métodos para classificar tráfego, que se foram modificando e evoluindo à medida que foram emergindo novas aplicações e serviços de crescente complexidade.

2.2.1 Classificação *Port-Based*

A IANA (*Internet Assigned Numbers Authority* [15]) definiu três categorias para os números dos portos: Portos *Well-Known*, Portos Registados e os Portos Privados e / ou Dinâmicos. A primeira categoria corresponde aos portos reservados para uso de aplicações e serviços específicos e têm números entre 0 e 1023. Os portos da segunda categoria têm números entre 1024 e 49151. Finalmente, os portos da última categoria estão localizados entre 49152 e 65535.

Este método quantifica o tráfego que utiliza portos *well-known* e é o método mais rápido e simples. Tendo em conta que as aplicações da Internet estão associadas a um porto, basta identificar que porto é usado pelo tráfego de dados analisado para identificar a aplicação em questão. Contudo, esta técnica tem bastantes limitações: muitas aplicações relativamente recentes tentam ultrapassar as *firewalls* e outros softwares de segurança disfarçando-se através da utilização de portos *well-known* associados a outras aplicações como máscara para os seus verdadeiros portos de operação; outras aplicações permitem que os utilizadores escolham manualmente os portos, alterando os portos existentes por defeito; a exaustão de endereços IPv4 disponíveis, obrigando servidores a disponibilizar serviços usando o mesmo endereço IP, mas através de diferentes portos. [16]

Esta técnica foi usada em alguns trabalhos ([17, 18]), obtendo-se taxas de classificação acertada inferiores a 70%. A. Madhukar et al. ([19]) verificaram que entre 30 a 70% do tráfego total analisado era tráfego “desconhecido”, ou seja, não era passível de ser classificado. Isto devia-se em grande parte aos programas *peer-to-peer* (P2P). Nos últimos anos tem havido vários problemas de ordem legal relativamente aos conteúdos disponibilizados por estes programas ([20]), o que tem levado os ISP a serem mais rigorosos com as possíveis infrações de propriedade intelectual. Assim, os programas P2P têm procurado dissimular a sua presença na rede, o que torna a classificação deste tráfego muito complexa. Estes programas podem corresponder a 80% do volume de tráfego total em certas regiões e horas do dia, daí a necessidade de classificar este tráfego. [21]

2.2.2 Classificação *Payload-Based*

Muitas aplicações e serviços têm uma “impressão digital” que facilita a sua identificação. Esta impressão digital consiste em sequências específicas de bytes presentes nos pacotes gerados por estas aplicações. Esta técnica de classificação consiste em inspecionar o conteúdo dos pacotes capturados para encontrar as assinaturas digitais das aplicações em causa. [19]

P. Haffner et al. ([22]) utilizaram três algoritmos de aprendizagem estatística de máquina no estudo: Naive Bayes, AdaBoost e Maxent (o Naive Bayes será abordado na secção 2.2.4). Estes algoritmos fizeram extração automática das assinaturas digitais de vários protocolos (FTP, SMTP, POP3, IMAP, HTTPS, HTTP e SSH) e verificou-se que obtiveram taxas de erro inferiores a 0.1%. Foi também atestada a durabilidade das assinaturas digitais destas aplicações, comparando capturas de tráfego espaçadas por sete meses e foi comprovado que a taxa de erro aumentou de forma mínima, comprovando que as assinaturas mantinham-se durante bastante tempo e os algoritmos tinham bom rendimento.

A. Madhukar et al. ([19]) utilizaram três técnicas diferentes de classificação: *Port-Based*, *Payload-Based* e Análise da Camada de transporte. Os resultados da primeira técnica já foram analisados na secção 2.2.1. Quanto à segunda técnica, os autores usaram a abordagem proposta por Sem et al. ([23]) para identificar o protocolo P2P, que consiste em usar as assinaturas digitais de cada aplicação (Tabela 2.1) e consultar documentação disponível. Os resultados obtidos foram bastante precisos e não houve ocorrências de tráfego não-P2P classificado como P2P.

Esta técnica tem algumas desvantagens: existem restrições de ordem legal e de privacidade à consulta dos conteúdos dos pacotes, vários protocolos usam encriptação no tráfego, faltam especificações e standards para as aplicações emergentes e ainda existe a eventualidade de diferentes implementações da mesma aplicação não seguirem as especificações oficiais. [14]

Protocolo P2P	Assinatura Digital
Gnutella2	“GNUTELLA”
KaZaA	“X-Kazaa”
BitTorrent	“.BitTorrent”

Tabela 2.1 – Protocolos P2P e respetivas assinaturas digitais (baseado em [19]).

2.2.3 Classificação *Host-Behavior Based*

Com esta técnica, é possível capturar a interação social observável de um *host*, mesmo que tenha conteúdo encriptável. Karagiannis et al. ([24]) analisaram os padrões de comportamento do *host* a três níveis (social, funcional e ao nível da aplicação) e foi conduzido sem acesso ao *payload* dos pacotes e sem conhecimento dos números dos portos usados. Ao nível social o comportamento do *host* é construído a partir da sua comunicação com outros *hosts*, medindo a sua popularidade conforme a quantidade de *hosts* com os quais contacta e identificando grupos de nós que podem constituir clientes com interesses em comum. Esta análise é efetuada apenas recorrendo aos endereços IP de origem e destino. Ao nível funcional, identifica-se o comportamento do *host* como fornecedor ou o utilizador de um determinado serviço (pode ser ambos, caso se trate de uma aplicação de âmbito colaborativo). Aqui o porto de origem é importante para determinar o papel do *host* a nível funcional, pois é provável que os *hosts* que usem apenas um porto para as suas comunicações com outros *hosts* sejam os fornecedores desse mesmo serviço. Ao nível da aplicação, são registadas as interações na camada de transporte ([13]) entre *hosts* em portos específicos, de forma a identificar a origem da aplicação. Os fluxos são 4-*tuple* (endereços IP e portos) e essa informação é depois complementada recorrendo a outras características como o tamanho dos pacotes. A

técnica usada neste estudo foi testada em três capturas diferentes de tráfego real e mais de 90% dos fluxos foram classificados com mais de 95% de exatidão.

A. Madhukar et al. ([19]) usaram análise estatística para classificar protocolos P2P, recorrendo a capturas sem tráfego UDP. Logo, toda a informação disponível provinha dos cabeçalhos de pacotes TCP, assim como das suas *flags* (TCP SYN, FIN e RST). Os resultados obtidos demonstram a existência de grande volume de tráfego P2P, validando portanto a fiabilidade desta técnica.

Tendo em conta que esta técnica faz a análise estatística de vários endereços IP, usando portos TCP e UDP e as *flags* TCP, não é aplicável a tráfego que circule em túneis seguros (como por exemplo os túneis IPsec [25]), pois mesmo que estes dados estejam disponíveis é provável que estejam relacionados com diferentes aplicações.

2.2.4 Classificação segundo as Características do Fluxo (*Flow-Features Based*)

Esta técnica classifica o tráfego tendo em conta a duração do fluxo, a quantidade e tamanho dos pacotes por fluxo, assim como o tempo de chegada entre pacotes.

Os algoritmos de classificação de máquina estão divididos em duas categorias: aprendizagem com supervisão e aprendizagem sem supervisão (também chamada de *clustering*). A primeira categoria precisa de tráfego de teste que já esteja classificado de modo a produzir um modelo que se ajusta a esse mesmo tráfego. Depois de "treinados", estes algoritmos conseguem detetar diferenças subtis entre fluxos de tráfego, de modo a classificar os fluxos de forma diferente. [26]

H. Kim et al. ([26]) compararam a performance de sete algoritmos de aprendizagem de máquina com supervisão (*Naive Bayes*, *Naive Bayes Kernel Estimation*, *Bayesian Network*, C4.5 *DecisionTree*, *k-Nearest Neighbors*, *Neural Networks* e *Support Vector Machines*/SVM) com os algoritmos de classificação BLINC ([24]) e Coral Reef ([27]), tendo em conta quatro condições: precisão, exatidão, memória e *F-Measure* (fórmula que compara a aplicação por fluxo, usando a memória e a precisão). Analisando a resposta dos vários algoritmos aos testes desenvolvidos pelos autores verifica-se que cada um deles apresenta respostas melhores a alguns parâmetros e que a solução ideal para melhorar cada algoritmo seria integrar aspetos de outros algoritmos onde ele mais falha. Contudo, o SVM foi o algoritmo com melhor performance geral (98% de exatidão), logo foi escolhido pelos autores ao desenvolverem um classificador que obtivesse bons resultados em vários cenários. Este classificador obteve taxas de exatidão entre 94% e 98%.

Erman et al. desenvolveram um estudo em que usaram dois algoritmos de classificação sem supervisão (K-Means e DBSCAN) para classificar tráfego e os resultados obtidos demonstraram que estes algoritmos são recomendados para agrupar tráfego com características e atributos semelhantes, mas não são capazes de etiquetar os clusters, tendo de recorrer à ajuda de outros algoritmos de classificação, nomeadamente os algoritmos de classificação com supervisão. [28]

Assim, conclui-se que tendo em conta as vantagens e desvantagens de cada um destes métodos de classificação de tráfego, é difícil dizer qual deles será o melhor: o método de classificação *Port-Based* é o mais rápido e simples, mas certas aplicações não podem ser classificadas pois utilizam certos portos como disfarce; o método *Payload-Based* é capaz de criar uma assinatura digital de uma aplicação analisando o seu conteúdo, o que por vezes pode levantar problemas ao nível da privacidade dos dados; o método *Host-Based* é bastante fiável na classificação de diferentes aplicações, mas não pode ser usado a tráfego de túneis seguros; finalmente, os algoritmos de aprendizagem supervisionada ou não supervisionada têm taxas de classificação certeira muito altas,

mas a eficácia em certos parâmetros é mais baixa que o desejável. No que concerne aos algoritmos de aprendizagem com supervisão, apresentam boas taxas de classificação acertada, mas não têm a mesma eficácia para todos os parâmetros considerados. Os algoritmos de aprendizagem sem supervisão têm a desvantagem de não serem capazes de classificar clusters, tendo de recorrer a algoritmos com supervisão para o poderem fazer.

2.3 Qualidade-de-serviço

O crescimento da Internet nos últimos anos assim como a evolução das tecnologias associadas à mesma permitiram uma maior oferta de serviços multimédia aos utilizadores, como por exemplo transmissão televisiva em direto, *vídeo on demand* ou o *streaming* AV. Contudo, estas aplicações multimédia requerem grandes recursos, como sejam a alocação de grande largura de banda e a assistência em tempo real [29]. Logo, o modelo de tráfego *best-effort* (todos os pacotes são tratados de forma igual) não se adequa a estas aplicações, devido aos requisitos anteriormente especificados, que tornam estas aplicações QdS-dependentes.

A QdS determina a utilidade e a facilidade de utilização de um determinado serviço ou aplicação, de modo a atestar a popularidade e funcionalidade do serviço em questão [30]. Assim, a QdS é um bom critério para fazer a distinção entre provedores de serviços semelhantes.

A QdS é definida por vários parâmetros que podem ser divididos em duas categorias: parâmetros humanos e parâmetros técnicos. Os parâmetros humanos, como o nome indica, são aqueles que dependem da intervenção direta humana enquanto os parâmetros técnicos dizem respeito às características da rede em si. Enumerando alguns parâmetros humanos, temos a estabilidade da qualidade do serviço, a disponibilidade das linhas subscritas, os tempos de espera, a informação ao utilizador ou a estabilidade de operação do sistema. Apesar dos requisitos de suporte de QdS variarem conforme a aplicação, existem alguns parâmetros importantes que nunca devem ser descurados, seja qual for a situação: [30]

- disponibilidade - probabilidade do serviço estar disponível e pronto a ser usado;
- acessibilidade - denota a taxa de sucesso de um determinado serviço atender um pedido. É possível que um serviço esteja disponível, mas não acessível;
- integridade – capacidade de um serviço manter uma interação correta e sem falhas com a fonte;
- largura de banda – um serviço que processe grandes quantidades de tráfego em pouco tempo precisa de mais espaço alocado em banda de modo a evitar falhas de comunicação;
- performance – um serviço tem uma performance alta quando atende muitos pedidos num certo período de tempo e uma latência baixa;
- confiança – é o parâmetro respeitante ao grau a que a aplicação é capaz de manter a estabilidade de serviço, assim como a sua qualidade;

- regulação - parâmetro segundo o qual cada serviço deve respeitar uma série de standards, regras, leis, assim como o acordo de nível de serviço estabelecido;
- segurança – o serviço tem de assegurar confidencialidade e autenticação de todas as partes envolvidas na ligação, encriptação de mensagens e controlo de acesso. Os níveis de segurança podem variar conforme o que for pedido ao prestador do serviço.

Conforme referido anteriormente, cada aplicação requer requisitos de rede com características e comportamentos próprios. Mapeando várias aplicações relevantes para este estudo, é possível perceber estas diferenças.

As aplicações multimédia (*streaming* AV, IPTV, VoIP, etc) combinam diferentes tipos de dados (gráficos, áudio e vídeo) que são normalmente processados de forma simultânea durante longos períodos de tempo (podem chegar a horas), o que torna estas aplicações extremamente sensíveis ao atraso. Logo, as aplicações multimédia requerem tempos de espera e latência o mais baixo possível, *jitter* (interferência no tempo de chegada de pacotes durante transferências de dados) atenuado e tempos médios de resposta baixos. Tendo em conta as grandes quantidades de tráfego geralmente envolvidas na execução destas aplicações estas requerem alocação constante de grande largura de banda em tempo real, dada a grande sensibilidade destas aplicações aos atrasos e interferências durante a transferência de dados [31].

A visualização de vídeos em sites como Youtube, Dailymotion ou Vimeo implica que o carregamento dos vídeos seja efetuado num curto espaço de tempo; quanto maior for o tempo de duração do vídeo, maior será o tempo de carregamento. Portanto, para um carregamento mais rápido é necessária a alocação de largura de banda constante, de modo a que o vídeo seja carregado rapidamente a partir da sua fonte. Mais uma vez, é importante que os tempos de espera e de resposta sejam pequenos. Atualmente muitos vídeos hospedados nestes sites estão disponíveis para visualização em diferentes definições (inclusivamente alta definição em muitos casos), portanto também é importante termos boas taxas de compressão e boas qualidades de som, cor e resolução de ecrã, ajustadas à qualidade pretendida pelo cliente.

Certas aplicações não têm requisitos de tempo real, como são os casos da consulta de e-mail, transferência de pequenos ficheiros, interação em redes sociais, programas de chat online e a visualização de notícias. Estas aplicações têm em comum o fato da navegação ser baseada em “cliques” do utilizador que lhe permite abrir novas páginas e aceder a diferentes conteúdos. Ao navegar, o utilizador tanto pode clicar repetidamente em relativamente pouco tempo como clicar poucas vezes num espaço de tempo maior; logo, nestes casos não é necessária a alocação de muita largura de banda, pois apenas surgem picos de utilização quando o utilizador tentar aceder a novas páginas. Comparativamente às aplicações multimédia, os requisitos de largura de banda são menores. Não havendo requisitos de tempo real, estas aplicações não apresentam tanta sensibilidade a atrasos como as aplicações multimédia ou de partilha de vídeos, o que leva a rede a responder de forma mais lenta (tempos de resposta e tempos de espera maiores).[29]

Existem vários mecanismos capazes de criar redes com QdS, com o objetivo de proporcionar um melhor serviço às aplicações nas orlas das mesmas [32]. Os mecanismos principais que providenciam QdS são o MPLS, RSVP e DiffServ.

2.3.1 MPLS/RSVP

O *Multi-Protocol Label Switching* (MPLS) é uma tecnologia de encaminhamento de pacotes para qualquer protocolo de rede. Integra informação da camada de Dados (largura de banda, utilização, latência, entre outros) na camada de Rede, facilitando a troca de pacotes ao nível de um sistema autónomo ou de um ISP [33]. Esta tecnologia utiliza endereços IP (tanto IPv4 como IPv6) para identificar pontos terminais e interfaces de *switches* e routers e encaminha os pacotes ao longo de percursos previamente definidos e configurados, os *Label Switched Paths* (LSP). Quando os pacotes entram numa rede baseada em MPLS, os routers limítrofes (*Label Edge Routers,* LER) atribuem um identificador aos pacotes de modo a distinguir os LSP; estes identificadores contêm dados do cabeçalho IP dos pacotes assim como informação relevante como o destino do pacote, largura de banda, tempo de espera ou o atraso. Depois de efetuada a associação de cada pacote a um LSP, os routers limítrofes enviam os pacotes pelo melhor caminho que é determinado pela consulta da tabela de encaminhamento de cada router. Enquanto os LER classificam cada pacote recorrendo a vários atributos, os *Label Switched Router* (LSR) presentes no interior da rede encaminham os pacotes recorrendo apenas aos identificadores LSP; ou seja, quando um LSR recebe um pacote, analisa o seu identificador para descobrir o LSP e encaminha o pacote pelo melhor caminho. Esta operação é repetida até que o pacote atinja um router que tenha de o enviar para fora da rede e aí é retirado o seu identificador. O MPLS é independente dos protocolos e pode transportar tráfego diverso e permite a gestores de rede e ISPs flexibilizar rotas na sua rede para evitar congestionamentos e falhas no tráfego. Do ponto de vista de QdS, o MPLS facilita a gestão e manipulação de fluxos de tráfego com diferentes requerimentos de alocação de recursos e diferentes prioridades. Por exemplo, clientes que tentam aceder a conteúdo que requer muita largura de banda e grande volume de tráfego constante podem ver reduzidos os atrasos e tempos de espera através desta tecnologia. [33, 34]

O *Resource Reservation Protocol* (RSVP) é um protocolo da camada de Transporte do modelo OSI ([13]) que permite a remetentes, recetores e routers envolvidos em sessões de comunicação (*multicast* e *unicast*) a possibilidade de comunicar entre si e reservar recursos necessários à transmissão de fluxos de dados *multicast* e *unicast*. O RSVP opera sobre IPv4 ou IPv6 e é usado pelos remetentes para pedir determinados parâmetros de QdS da rede para fluxos de dados de aplicações particulares e pelos routers para entregar pedidos de parâmetros QdS para todos os nós ao longo do percurso dos fluxos. [35]

O protocolo RSVP tem como principais atributos: a possibilidade de suportar fluxos de dados *multicast* e *unicast*; capacidade de fazer reservas para fluxos de dados unidirecionais; é orientado ao recetor, pois é este que inicia a reserva de recursos usados para o fluxo que recebe; mantém *soft-state*, ou seja, as reservas e alterações nos fluxos são realizadas de acordo com condições estáticas e dinâmicas; depende dos protocolos de *routing* existentes e futuros, apesar de não ser um protocolo de *routing*; transporta QdS, tráfego e parâmetros de controlo de igual forma, pois não sabe o conteúdo dos objetos que transporta. [35, 36]

A Figura 2.2 representa um modelo de funcionamento do RSVP numa sessão *multicast* com um remetente de tráfego (S1) e três recetores (RCV1, RCV2 e RCV3). As primeiras mensagens enviadas por RSVP são as mensagens *Path* originárias de S1 e as mensagens *Resv* oriundas de cada um dos recetores. As mensagens *Path* têm como função instalar em cada router o estado do pedido de reserva pendente (*Path state*) e assegurar que os recetores têm toda a informação sobre o tráfego enviado por S1 e o percurso entre ambos os extremos, para que os recetores possam fazer os pedidos de reserva. Já as mensagens *Resv* têm a tarefa de transportar pedidos de reserva até aos routers ao longo do percurso entre recetores e o remetente.

A partir do momento em que existem dados com QdS em S1 para serem enviados, este começa a emitir periodicamente mensagens *Path* para o primeiro router, R1. Cada router verifica se as mensagens *Path* têm erros: em caso afirmativo, o router envia uma mensagem *Path Err upstream* para informar o remetente da falha na mensagem; caso a mensagem seja válida, o router atualiza o *Path State*, incluindo o endereço *Phop* do anterior router *upstream*, de modo a que as mensagens *Resv* possam seguir o percurso *upstream.* Cada router é também responsável por enviar as mensagens ao longo do percurso.

Quando os routers recebem mensagens *Resv* dos recetores fazem uma reserva dos parâmetros requisitados e enviam o pedido na direção de S1, se o controlo de tráfego existente em cada router determinar que existem recursos para satisfazer o pedido. Cada mensagem *Resv* contém os parâmetros QdS requisitados, assim como os pacotes a que se destinam tais parâmetros, logo esta informação é acedida em cada router. Quando as mensagens *Resv* chegam com sucesso ao remetente e estando todos os parâmetros de acordo com o reservado pelo(s) recetor(es), o remetente procede ao envio de dados. Toda esta operação é realizada em *soft-state*, portanto se durante algum tempo não existir circulação de informação, o pedido de reserva é cancelado. [36, 37]

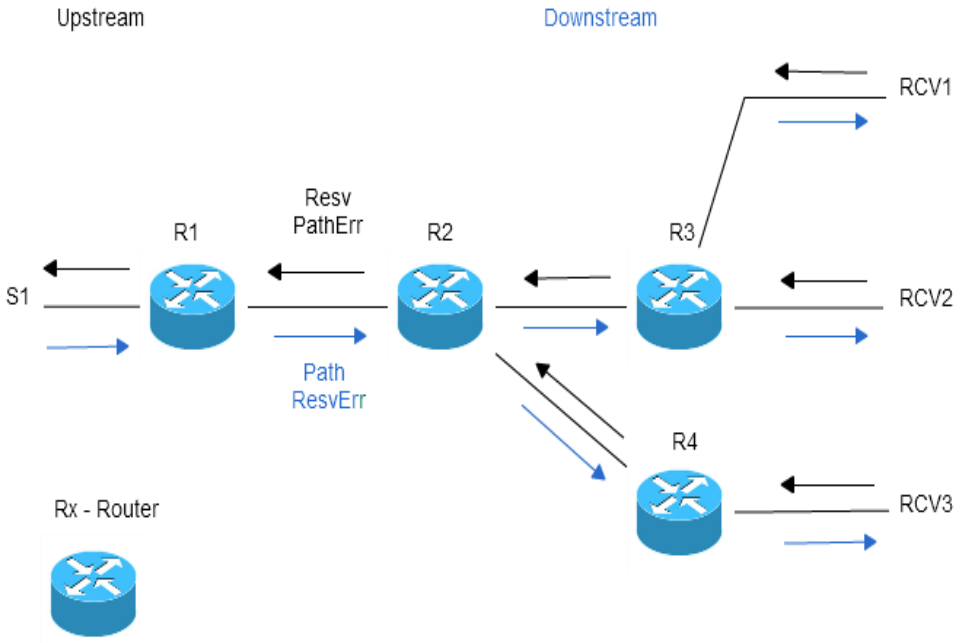

Figura 2.2 - Esquema do modelo RSVP com a direção das mensagens RSVP (*Path* e *Resv*). (Figura baseada em [35]).

2.3.2 DiffServ

O modelo DiffServ (*Differentiated Services*) é uma arquitetura de rede que providencia parâmetros de QdS e gere tráfego em redes de larga escala, onde podem coexistir milhares de sessões em simultâneo. No modelo DiffServ, o tráfego que entra na rede é classificado e condicionado nos nós limítrofes da rede. Este condicionamento envolve algumas ações sobre o tráfego como medição, modelação e marcação. Cada pacote é classificado individualmente: no cabeçalho IP de cada pacote é definido um campo de 6 bits, chamado *DiffServ Codepoint* (DSCP), que indica a que agregado de

fluxos pertence o pacote em questão. Assim, os remetentes que enviem tráfego numa rede DiffServ marcam cada pacote com um valor DSCP. Tráfego de diferentes fluxos que tenha parâmetros de QdS semelhantes é marcado com o mesmo DSCP, agregando assim os fluxos com comportamento semelhante. [38, 39]

No core da rede, é o *per-hop behavior* (PHB) que determina como cada nó distribui os recursos para os diferentes agregados de comportamento. PHB são especificados em termos da prioridade relativa de recursos (buffers, largura de banda) em relação a outros PHB ou em termos das características do seu tráfego (atrasos e perdas de pacotes). Nestes casos, estes PHB são especificados como um grupo e devem ser usados como um bloco, de forma a alocar recursos de forma consistente e organizada. [39]

O modelo de QdS DiffServ oferece uma maior escabilidade que o modelo IntServ / RSVP, pois efetua reservas para agregados de fluxos e não para fluxos individuais.

3 Fundamentação Teórica da Análise Multi-Escalar

Antes da descrição da importância das *wavelets* e escalogramas na análise multi-escalar de tráfego, é importante abordar as Transformadas de Fourier (FT). As Transformadas de Fourier são especialmente adequadas para a análise do espetro de frequência de processos estacionários (processos com a mesma componente de frequência em toda a sua gama), decompondo-os em diferentes componentes (funções exponenciais complexas), sendo que cada um destes componentes está associado a uma frequência própria [40]. Contudo, as FT não são capazes de analisar espetros de frequência de processos não-estacionários, pois estes requerem uma representação relativa ao tempo e à frequência e as FT apenas são capazes de trabalhar com sinais que tenham apenas uma componente de frequência e que não varie com o tempo. As *wavelets* resolvem este problema.

3.1 *Wavelets* e Escalogramas

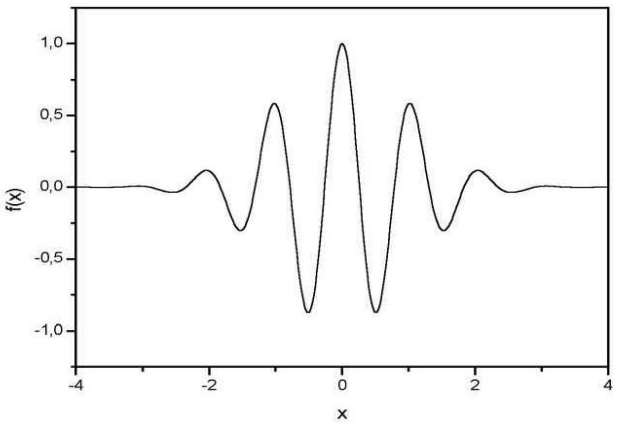

Figura 3.1 – Representação de uma *wavelet Morlet*.

Wavelets são funções matemáticas que quando aplicadas a um certo sinal, dividem-no em diferentes componentes de frequência. As *wavelets* são ondas de curta duração com energia limitada, como é observável na Figura 3.1, e possibilitam a análise individual de cada componente do sinal numa escala apropriada [41]. Tendo em conta a sua curta duração, as *wavelets* oferecem boa resolução de tempo e frequência, o que constitui uma vantagem em relação às FT que apenas analisam informação de sinais com a mesma componente de frequência.

Scaling pode definir-se como uma das propriedades da invariância de escala que se pode encontrar nas redes de tráfego; um exemplo de *scaling* é a natureza auto-similar (o objeto como um todo tem a mesma forma de uma ou mais das partes que o constituem) das mesmas. [14, 42]

Uma *wavelet* $\psi(t)$ pode ser definida como uma função passa-banda que oscila em torno de uma frequência central, satisfazendo assim a seguinte condição:

$$0 < C_\Psi = 2\pi \int_{-\infty}^{+\infty} \frac{|\Psi(w)|^2}{|w|} dw < \infty, \quad (3.1)$$

sendo que C_Ψ é o fator de admissibilidade e $\Psi(w)$ é a FT de ψ. Para obter esta condição, basta que se cumpra:

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \tag{3.2}$$

Esta condição assegura que a média desta função desaparece, o que implica que as *wavelets* tenham a forma de uma onda e sejam definidas como uma função passa-banda. [43, 44]

Assumindo uma *wavelet* "mãe" $\psi(t)$, ao efetuar-se o *scaling* e a translação desta *wavelet* "mãe" obtém-se uma família de "*wavelets* filhas" $\psi_{\tau,s}(t)$:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\psi(\frac{l-\tau}{s}), \tag{3.3}$$

em que s é um fator de *scaling* ou dilatação que controla a largura da janela de análise da *wavelet*. O parâmetro $\frac{1}{\sqrt{|s|}}$ garante a preservação da energia ($\left\|\psi_{\tau,s}\right\| = |\psi|$) e τ é um parâmetro relativo à translação que controla a movimentação da *wavelet*. A translação da *wavelet* é obtida através da deslocação da sua posição ao longo do tempo. O *scaling* da *wavelet* varia conforme o valor do módulo de s: se$|s| > 1$, a wavelet é alargada; se$|s| < 1$, a *wavelet* é comprimida [41, 44]. Ambos os parâmetros são obrigatoriamente incrementados continuamente, pois esta é uma transformada contínua.

Dada uma série temporal $x(t) \in L^2(IR)$, a CWT (*Continuous Wavelet Transform*) relativa à *wavelet* "mãe" ψ é uma função de duas variáveis:

$$\psi_{x,\psi}(\tau, s) = < x, \psi_{\tau,s} > \int_{+\infty}^{-\infty} x(t)\frac{1}{\sqrt{|s|}}\psi * \left(\frac{l-\tau}{s}\right)dt, \tag{3.4}$$

em que * é o complexo conjugado de $\psi(t) \in L^2(IR)$. Analogamente ao caso das séries de Fourier referido anteriormente, o espetro de potência da *wavelet* (também conhecido como escalograma, como irá ser referenciado a partir de agora) é dado por:

$$\hat{S}_x(\tau, s) = 100\frac{|W_x(\tau,s)|^2}{\sum_{\tau'}\sum_{s'}|W_x(\tau,s)|^2} \tag{3.5}$$

A equação 3.5 está ilustrada na Figura 3.2, no gráfico inferior. Analisando estes escalogramas, é possível avaliar a existência das diferentes componentes de frequência da série temporal em questão. A ocorrência de picos na região das baixas frequências do escalograma indica a existência de um componente de baixa frequência na série temporal. Analogamente, a ocorrência de picos na região das altas frequências do escalograma indica a existência de um componente de alta frequência na série temporal. Os escalogramas requerem elevados recursos computacionais, sendo portanto recomendados para tarefas *offline* e são utilizados em várias áreas científicas devido à sua capacidade de análise de processos não-estacionários [14, 41]. A CWT vai ser usada no âmbito desta dissertação pelas suas qualidades na disponibilização de informação em termos de frequência e do tempo.

É possível efetuar classificação de tráfego recorrendo a outro tipo de *wavelet*, a DWT (*Discrete Wavelet Transform*). A DWT também representa funções com resolução de tempo e frequência. Neste caso, a decomposição da *wavelet* utiliza uma função

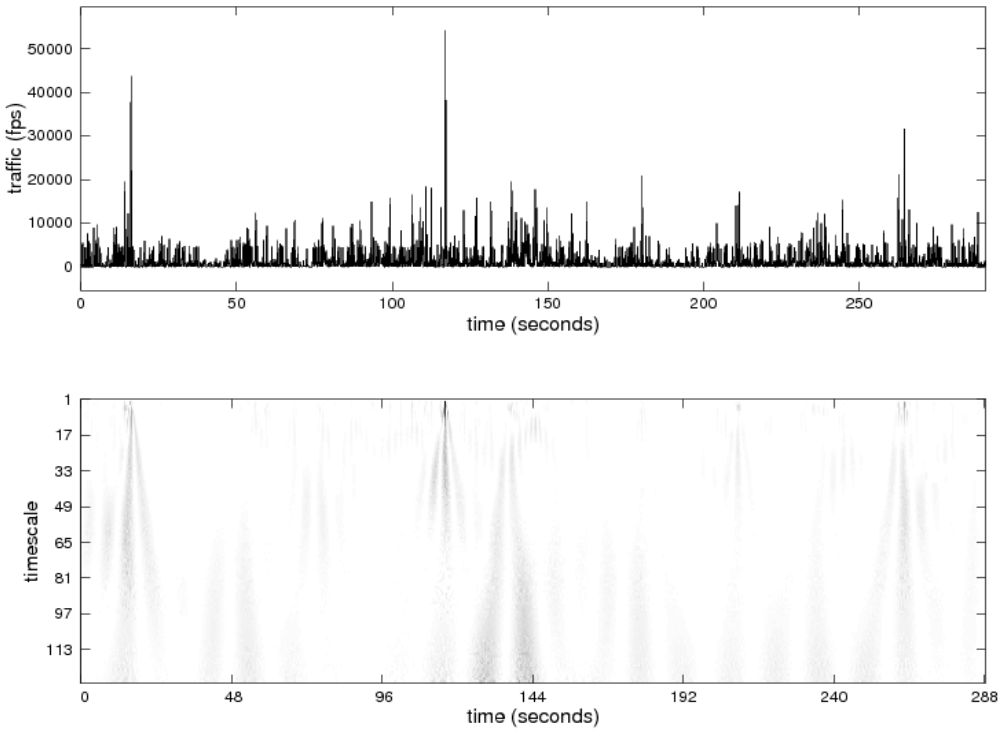

Figura 3.2 - Exemplo de Escalograma.

passa-baixo, que efetua o *scaling* e que pode deslocar-se temporariamente. Assim, a equação que representa a DWT é dada por:

$$X(t) = \sum_k c_x(j_0,k)\phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_k d_X(j,k)\,\psi_{j,k}(t) \quad (3.6)$$

em que $c_x(j_0,k)$ são os coeficientes do scaling e $d_X(j,k)$ são os coeficientes da *wavelet.* [43, 44]

3.2 Diferenciação Multi-Escalar

No âmbito desta dissertação, a análise multi-escalar é efetuada recorrendo a escalogramas (CWT) que decompõem em várias escalas o tráfego em questão segundo as suas componentes de tempo e frequência; dessa decomposição resulta um espetro que congrega as várias componentes de frequência. Os perfis dos utilizadores são construídos tendo em conta a divisão desse espetro em várias regiões com diferentes componentes de frequência e posterior correlação desses componentes. [3]

O tráfego na Internet é gerado a partir de pedidos dos utilizadores (geralmente eventos de baixa frequência) e controlado por mecanismos de controlo de tráfego (eventos de frequência média). A resposta a esses pedidos chega na forma de pacotes que correspondem aos eventos de frequência alta. O simples ato de clicar num link aciona uma cadeia de eventos (ilustrados na Figura 3.3): o sistema operativo cria vários processos; cada um destes processos concebe uma lista de sessões Internet; cada sessão origina um fluxo de tráfego. Ao nível da camada de Rede [13], cada um destes fluxos vai transmitir e receber os dados pretendidos pelo utilizador em vários pacotes [3]. A Figura 3.3 demonstra como as diferentes escalas (tempo /frequência) se interligam de modo a moldar os fluxos de tráfego.

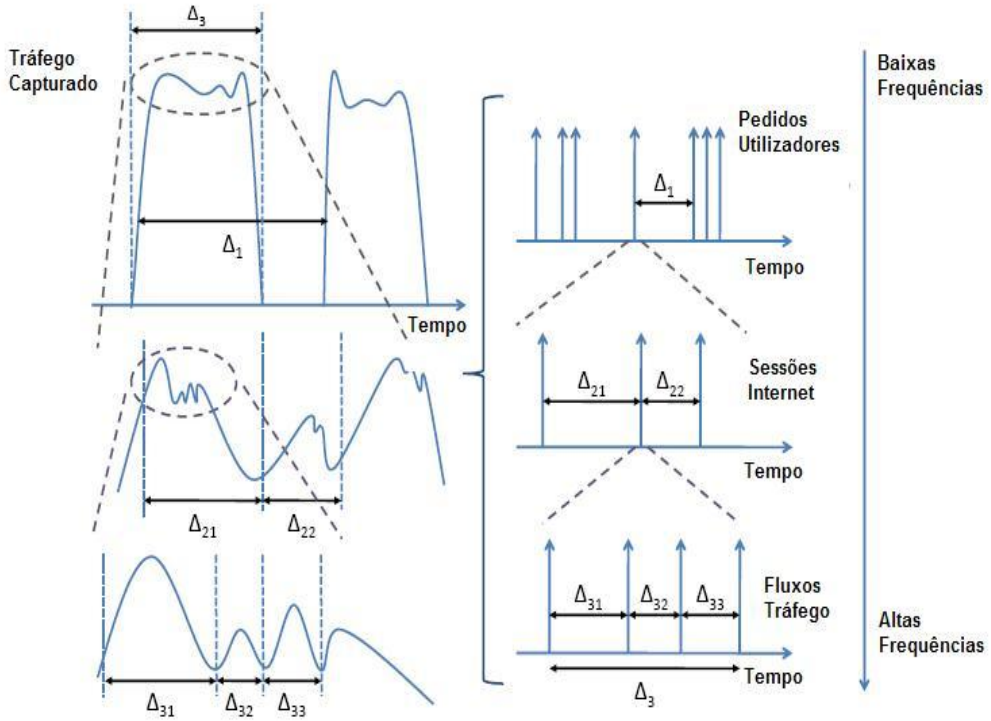

Figura 3.3 – Dinâmica de tráfego multi-escalar: Δ_1 –intervalo de tempo entre pedidos dos utilizadores; Δ_{2x} –instante de início da transmissão de pacotes; Δ_{3x} –instante de chegada dos pacotes (Figura baseada em [3])

A Figura 3.4 relaciona as componentes de frequência de cada tipo de evento com a energia associada ao mesmo. A região das baixas frequências está relacionada com as ações dos utilizadores humanos e com os *scans* efetuados por ordem do sistema de rede. Na região das médias frequências inserem-se principalmente eventos normalmente gerados pela rede (criação de sessões, controlo e modelação de tráfego, por exemplo) em resposta a pedidos de dados que recebe por parte dos utilizadores. Na região das altas frequências inserem-se os eventos que envolvem elevado tráfego de pacotes. Como exemplos deste tipo de tráfego temos a visualização de vídeos online, a consulta de websites com vídeos embutidos (caso dos sites de notícias) ou ainda a transferência de ficheiros por email. Estes são eventos facilmente identificáveis no espetro de frequência, pois o carregamento de vídeos e a transferência de ficheiros estão normalmente associados a um grande desvio-padrão de energia. A criação de sessões para transferência de ficheiros e a ordem de execução de *scans* também são evento com grande desvio-padrão de energia. No pólo oposto, a performance do *scan* é um evento com poucos componentes de frequências médias, pois envolve poucos fluxos de tráfego que por sua vez não apresentam muita variação. [3]

Este mapeamento constitui uma ferramenta preciosa na classificação das aplicações de Internet, pois sabendo os tipos de eventos que cada fluxo de tráfego gera, torna-se mais fácil fazer uma relação eficaz entre as aplicações e os fluxos de tráfego.

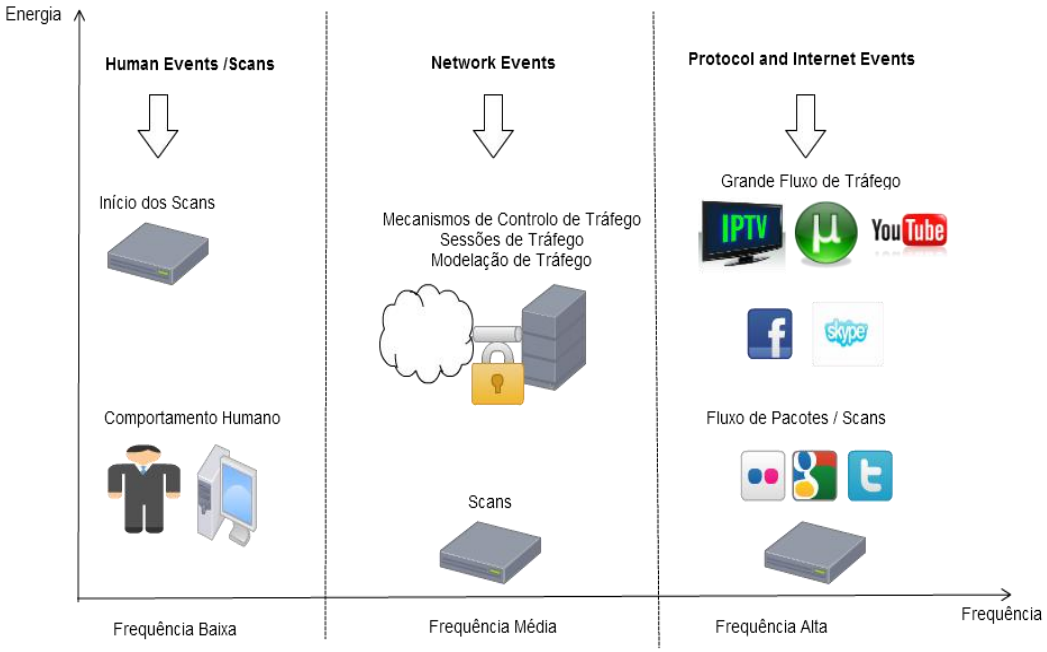

Figura 3.4 - Mapeamento dos mecanismos de rede e de utilizadores tendo em conta as regiões de variação das frequências (Figura baseada em [3]).

4 Recolha e Processamento do Tráfego

Todas as medições e testes necessários no âmbito desta dissertação foram efetuados no Laboratório de Redes 1 do Instituto de Telecomunicações na Universidade de Aveiro. Os testes foram executados recorrendo a um computador com processador Intel Core i5 CPU 650 @ 3.20 GHz x 4. Foi utilizado o sistema operativo Ubuntu 12.04 LTS.

4.1 Capturas de Tráfego

As capturas de tráfego utilizadas para a execução dos testes foram obtidas através do CAIDA (*The Cooperative Association for Internet Data Analysis*). As capturas disponibilizadas pelo CAIDA têm a duração de uma hora e são obtidas uma vez por mês em cada um dos seus monitores passivos. [45, 46]

Os conjuntos de dados (*datasets*) utilizados nesta dissertação foram recolhidos no dia 21 de Julho de 2011 a partir das 12h59 UTC (*Coordinated Universal Time*) com a duração de uma hora, no *datacenter* Equinix situado em Chicago (Illinois, EUA), que está conectado à linha *backbone* de um ISP de nível 1 entre Chicago e Seattle (Washington, EUA). Esta ligação é bidirecional, portanto o tráfego de pacotes que circula nas duas capturas de tráfego utilizadas tem direções opostas: na captura de tráfego da direção A, o tráfego de pacotes tem origem em Seattle e destino em Chicago; na captura de tráfego da direção B, os pacotes fluem no sentido contrário, de Chicago para Seattle [47].Depois da captura, o CAIDA procede à anonamização do tráfego de forma diferenciada para cada direção e impõe ainda uma política de uso restrito por questões de ordem legal e de privacidade.

Estas capturas de tráfego estão divididas em ficheiros ".pcap", cuja duração varia entre os cinquenta e os cinquenta e nove segundos. Tendo em conta o âmbito desta dissertação, foram selecionados os primeiros cinco minutos das capturas de tráfego de ambas as direções para serem efetuados os respetivos testes. Os fluxos de ambas as capturas de tráfego são identificados por um *5-tuple* (endereço IP de origem (srcIP), endereço IP de destino (dstIP), porto de entrada (srcPrt), porto de saída (dstPrt) e protocolo). Convém realçar que existe uma diferença no fuso horário entre as duas cidades: no início da captura do tráfego, a hora local em Chicago era 6h59 UTC-6h00 e a hora local em Seattle era 4h59 UTC-8h00 [48]. Esta diferença de fuso horário entre as duas cidades ajuda a explicar o desnível no fluxo de tráfego nas duas direções.

De seguida apresentam-se características importantes de alguns protocolos mais relevantes nas capturas de tráfego analisadas e que serão alvo de estudo ao longo deste trabalho, pois englobam diferentes classes de serviço e várias aplicações com crescente importância na internet atual. Assume-se que os pacotes com tamanho superior a 256 bytes presentes nos fluxos em análise serão pacotes de dados enquanto os restantes serão pacotes de controlo.

4.1.1 HTTP

O HTTP (*Hypertext Transfer Protocol)* é um protocolo da camada de Aplicação [13] e um dos pilares da *World Wide Web*. Este protocolo é uma sequência de interações pedido-resposta que permite a comunicação entre servidores e clientes. Considerando um browser como o cliente e uma aplicação ativa num computador que hospeda um

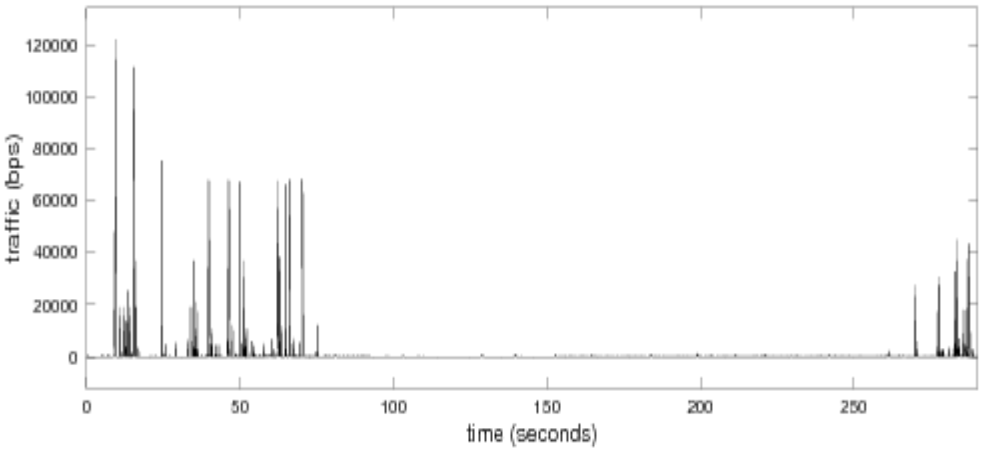

Figura 4.1 – Tráfego *downstream* HTTP por parte do cliente na direção A (bytes por segundo).

website como o servidor, a sessão inicia quando o cliente estabelece uma ligação TCP (*Transmission Control Protocol*) para um porto no servidor (geralmente o destino é o porto 80, porto atribuído por defeito ao HTTP). De seguida, o cliente envia um *HTTP request* (pedido) para o servidor com a identificação do(s) ficheiro(s) ao qual(ais) pretende aceder; o servidor verifica o conteúdo do pedido e envia um *HTTP response* para o cliente com uma atualização do estado e uma resposta negativa ou positiva, sendo que neste caso envia também o conteúdo do(s) ficheiro(s) pretendido(s).

Na camada de Transporte, o protocolo HTTP trabalha predominantemente com o protocolo TCP, embora o UDP (*User Datagram Protocol*) também possa ser eventualmente usado para transporte de pacotes de certas aplicações (contudo esta situação não é tão frequente).

Seguidamente apresentam-se alguns exemplos de tráfego HTTP capturado nos *datasets* estudados.

A Figura 4.1 é o exemplo típico de tráfego *downstream* HTTP por parte do utilizador com picos não periódicos de curta duração, mas de grande amplitude. Estes picos correspondem aos cliques do utilizador quando este acede a uma nova página. Existem momentos onde o tráfego é nulo, pois correspondem aos instantes em que o utilizador está a visualizar a página que abriu; estes instantes são particularmente visíveis entre o primeiro minuto e o último minuto, onde o utilizador volta a abrir novas páginas. [41, 44]

A Figura 4.2 apresenta um grafismo semelhante ao normalmente apresentado quando um utilizador acede a redes sociais. Existem poucos picos de tráfego, logo o utilizador não está a abrir novas páginas com grande regularidade e estas páginas não apresentam conteúdos muito pesados. Verifica-se que o utilizador não demora muito tempo a visualizar as páginas, pois existem poucos períodos sem ocorrência de tráfego. [41]

A Figura 4.3 tem características normalmente associadas à visualização de websites de partilha de fotos. O tráfego é periódico e é possível verificar que o download dos conteúdos requer alguns instantes, o que faz sentido se esse conteúdo for imagens hospedadas online. Além disso, os intervalos de tempo em que o tráfego é reduzido ou nulo são extremamente curtos, o que se coaduna com a visualização de imagens. [41]

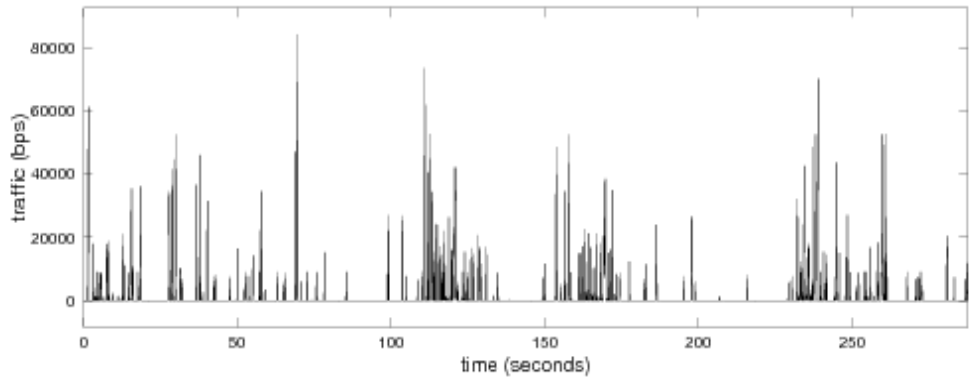

Figura 4.2 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

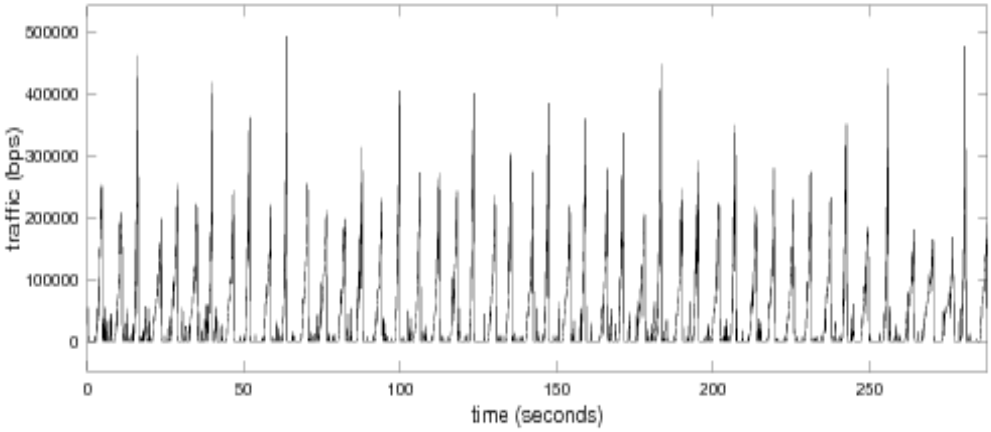

Figura 4.3 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

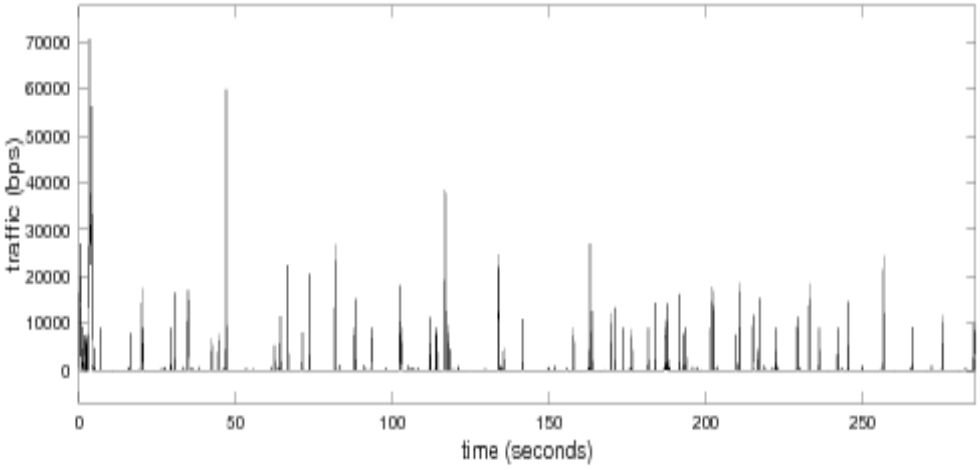

Figura 4.4 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

A Figura 4.4 apresenta tráfego característico do *browsing*: picos de tráfego pouco frequentes, irregulares temporalmente e períodos de tempo de alguns segundos sem qualquer tráfego, o que pode indicar que o utilizador esteja a ler informação. Os pacotes de menor tamanho podem indiciar a presença de alguma publicidade e *pop-ups* nas páginas. [44]

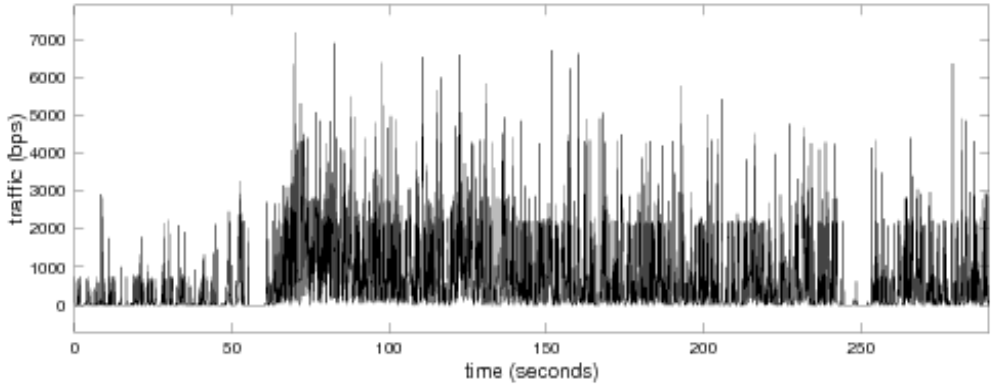

Figura 4.5 - Tráfego *downstream* HTTP por parte do cliente na direção A (bytes por segundo).

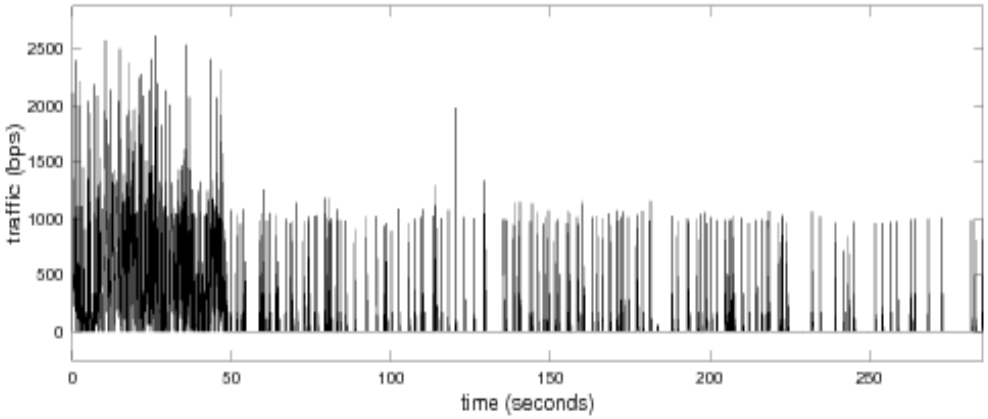

Figura 4.6 - Tráfego *upstream* HTTP por parte do cliente na direção A (bytes por segundo).

A Figura 4.5representa o exemplo de tráfego com abertura regular de páginas, pois verifica-se um grande volume de pacotes durante um período de tempo bastante alargado (neste caso sensivelmente três minutos), com picos de tráfego separados por poucos segundos, o que indicia que o utilizador estará a fazer *browsing* entre páginas.

No que diz respeito ao tráfego *upstream* HTTP, o tamanho dos pacotes é substancialmente menor comparando com o cenário do tráfego *downstream* HTTP abordado até ao momento. Na Figura 4.6, verifica-se um grande volume de picos de tráfego durante os primeiros cinquenta segundos, de forma praticamente contínua, o que indicia que o utilizador está extremamente ativo. A partir daí, os picos de tráfego deixam de ser frequentes, o que alude a uma redução da sua atividade.

No caso da Figura 4.7, existem ocasionalmente picos de tráfego, principalmente nos dois primeiros minutos. O restante fluxo compreende pacotes de poucas dimensões. A atividade do utilizador reduz-se de forma evidente a partir do terceiro minuto.

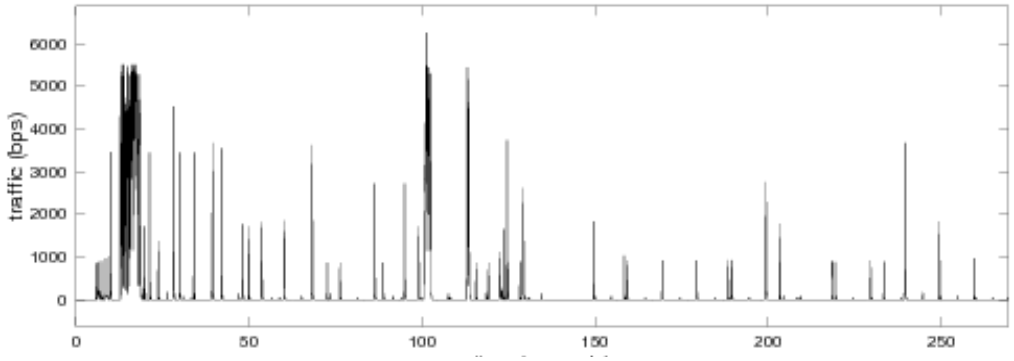

Figura 4.7 - Tráfego *upstream* HTTP por parte do cliente na direção B (bytes por segundo).

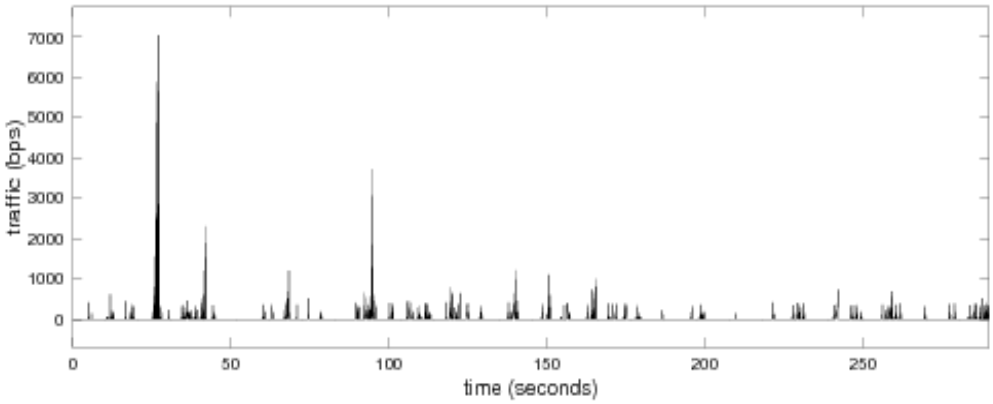

Figura 4.8 - Tráfego *upstream* HTTP por parte do cliente na direção A (bytes por segundo).

Relativamente à Figura 4.8, verifica-se que existe apenas um pico de tráfego, o que pode significar que o utilizador fez poucos cliques. A chegada regular de pacotes de menor tamanho indicia que o utilizador continua a visualizar páginas ao longo dos cinco minutos, embora existam intervalos de tempo de alguns segundos sem qualquer tipo de tráfego.

4.1.2 SMTP

O SMTP (*Simple Mail Transfer Protocol*) é um protocolo da camada de Aplicação que permite a transmissão de correio eletrónico ao longo da rede, embora seja predominantemente usado pelos clientes de email apenas para entrega de mensagens. O porto TCP atribuído a este protocolo é o porto 25.

Uma sessão SMTP consiste numa sequência de comandos enviados pelo cliente SMTP e as respetivas respostas do servidor SMTP de modo a criar a sessão e trocar parâmetros relacionados com a mesma. O cliente SMTP pode ser a aplicação de gestão de email de um utilizador (*mail user agent*, MUA) ou então um servidor SMTP a atuar como cliente SMTP de modo a poder retransmitir emails (*mail transfer agent*, MTA). [49]

De seguida apresentam-se alguns exemplos de tráfego SMTP capturado nas capturas de tráfego utilizadas.

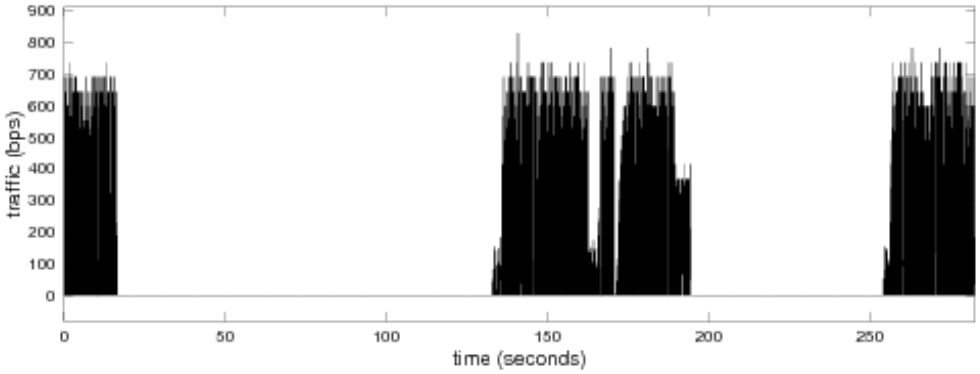

Figura 4.9 – Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

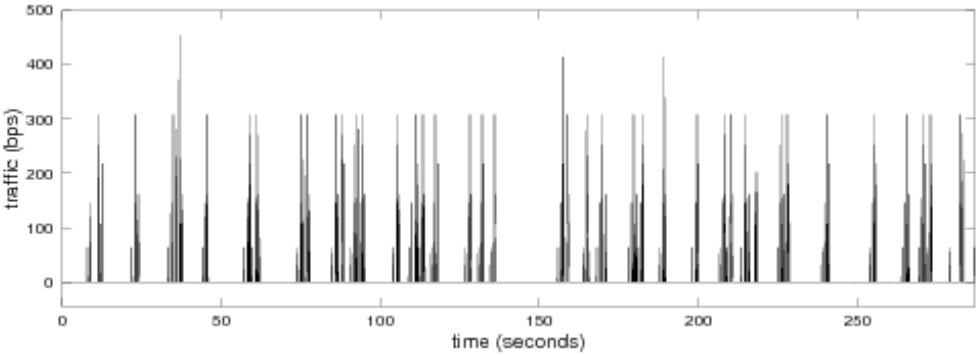

Figura 4.10 - Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

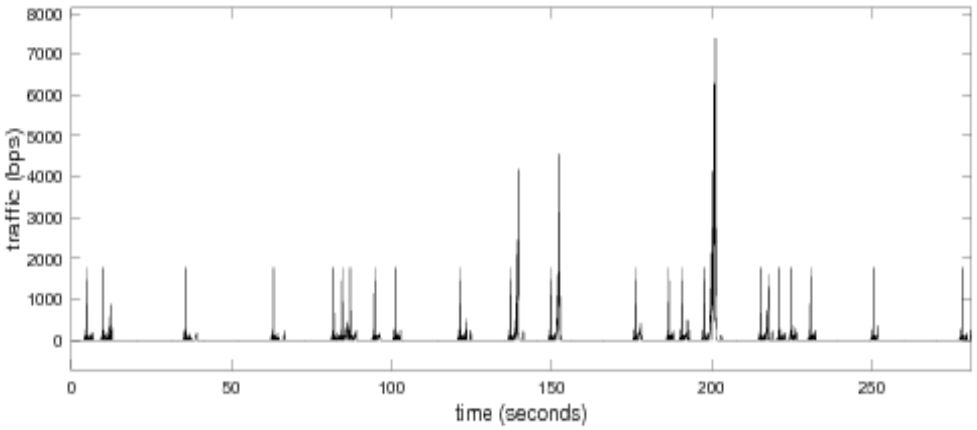

Figura 4.11 - Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

Na Figura 4.9 verifica-se que existem três grandes concentrações de tráfego em curtos períodos de tempo: sensivelmente nos dez primeiros segundos da amostra; durante grande parte do segundo minuto e finalmente nos últimos segundos. Tendo em conta a natureza deste protocolo, pode assumir-se que os pacotes capturados referem-se ao acesso do utilizador à sua caixa de correio para enviar emails.

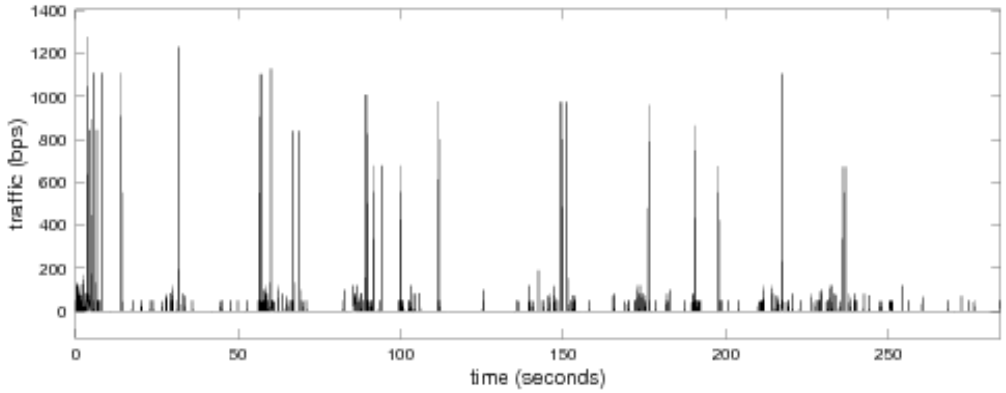

Figura 4.12 - Tráfego *upstream* SMTP por parte do cliente na direção A (bytes por segundo).

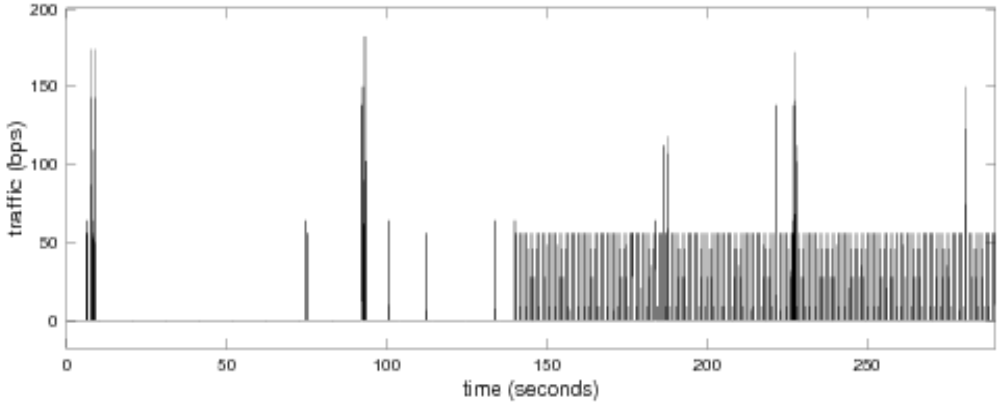

Figura 4.13 - Tráfego *upstream* SMTP por parte do cliente na direção A (bytes por segundo).

Analisando a Figura 4.10, verifica-se que o tráfego é periódico e existem poucos picos de tráfego, o que aponta para pouca atividade do utilizador. Este pode ser um caso em que o utilizador faz *refresh* (atualiza)à sua caixa de correio para verificar se recebeu novos emails.

O caso da Figura 4.11 apresenta poucos picos de tráfego, que podem estar relacionados com o envio de um email e respetivo tráfego de controlo dessa transferência. O restante tráfego pode dever-se a atualizações da caixa de correio do utilizador.

No que diz respeito ao tráfego *upstream* SMTP, a existência de picos de tráfego bastante espaçados no tempo na Figura 4.12 pode indiciar a criação de múltiplas sessões SMTP por parte do cliente. Nesta amostra de cinco minutos o restante tráfego, embora de pacotes de controlo, surge normalmente associado a estes picos. Existem também intervalos de tempo sem qualquer tipo de atividade ou atualização.

Analisando a Figura 4.13, verifica-se que existem poucos picos de tráfego, mas bem identificados, e a partir do terceiro minuto da amostra o tráfego torna-se constante. Tendo em conta o tamanho dos pacotes encontrados, este tráfego é sobretudo de controlo, não havendo portanto envio de emails.

4.1.3 POP3

O POP (*Post Office Protocol*) é um protocolo da camada de Aplicação, utilizado por clientes de email para receberem emails, seja através de servidores ou via uma ligação TCP. O POP foi desenvolvido ao longo dos anos, sendo a sua última versão (e mais utilizada) o POP3. O porto TCP atribuído a esta aplicação é o porto 110.

O POP3 funciona como interface entre as aplicações de cliente de email e as unidades de armazenamento dos mesmos: monitoriza o porto 110 para receber ligações de aplicações de email, autentica a aplicação e por fim faz a gestão da entrega da mensagem com essas mesma aplicação. O processo de autenticação é efetuado recorrendo a um repositório onde estão armazenadas todas as informações necessárias acerca dos utilizadores. [50]

O POP3 complementa-se com o SMTP, pois o serviço de entrega de mensagens do SMTP trabalha em conjunto com o protocolo POP3 assim que a mensagem enviada chega ao ISP do destinatário. O serviço de entrega do SMTP é notificado pelo serviço SMTP assim que um novo email chega; de seguida o POP3 executa os comandos apresentados anteriormente, encaminhando o novo email para a unidade de armazenamento de emails. [50]

Apresentam-se agora exemplos de tráfego POP3 existente nas capturas de tráfego utilizadas. As características da Figura 4.14 assemelham-se às do download de um ficheiro, devido ao tráfego constante de pacotes longos durante um intervalo de tempo bastante alargado, com muitos picos de tráfego. Tendo em conta o protocolo em causa, pode assumir-se que neste exemplo está a ser efetuado o download de um email com um ou mais anexos de dimensão relativamente grande.

Na Figura 4.15 atesta-se a existência de tráfego relevante durante alguns segundos com apenas um pico de tráfego, o que pode indiciar a receção de um email. Os pacotes recolhidos durante o restante intervalo da amostra são de tamanho pequeno e praticamente impercetíveis nesta figura.

A Figura 4.16 apresenta tráfego periódico e com pacotes de tamanho semelhante ao longo de toda a janela temporal da amostra, sem picos de tráfego. Isto significa que este tráfego não está associado à receção de um ou mais emails, mas pode ser tráfego de controlo de atualização da caixa de correio do utilizador.

Na Figura 4.17, verifica-se a criação de várias sessões POP3 com duração temporal diversificada, mas sempre na casa das dezenas de segundos. Grande parte dos pacotes tem pouco comprimento, sendo portanto maioritariamente tráfego de controlo. Existe apenas um pico de tráfego claro, o que pode indiciar atividade breve do utilizador.

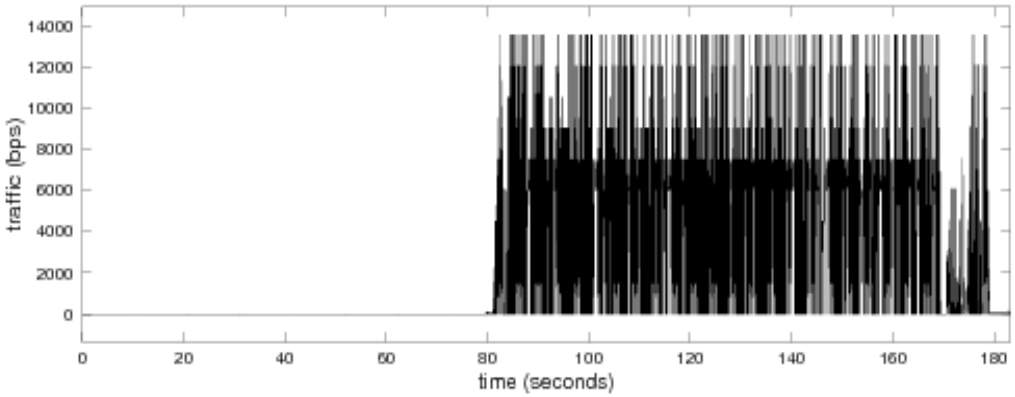

Figura 4.14 - Tráfego *downstream* POP3 por parte do cliente na direção A (bytes por segundo).

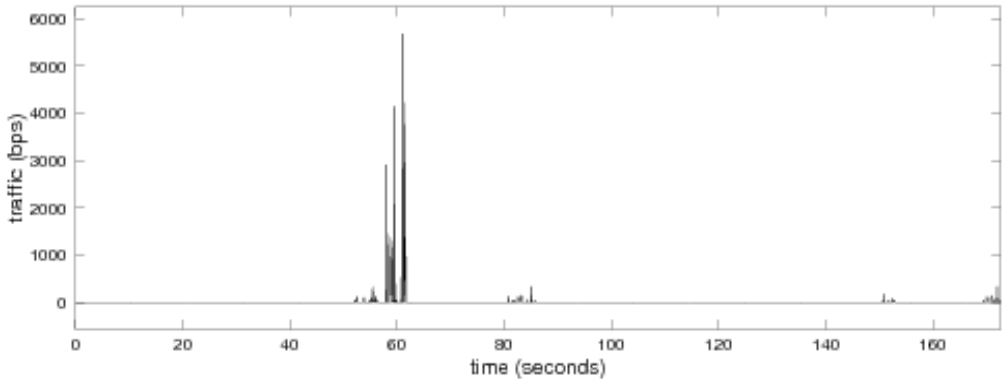

Figura 4.15 -Tráfego *downstream* POP3 por parte do cliente na direção A (bytes por segundo).

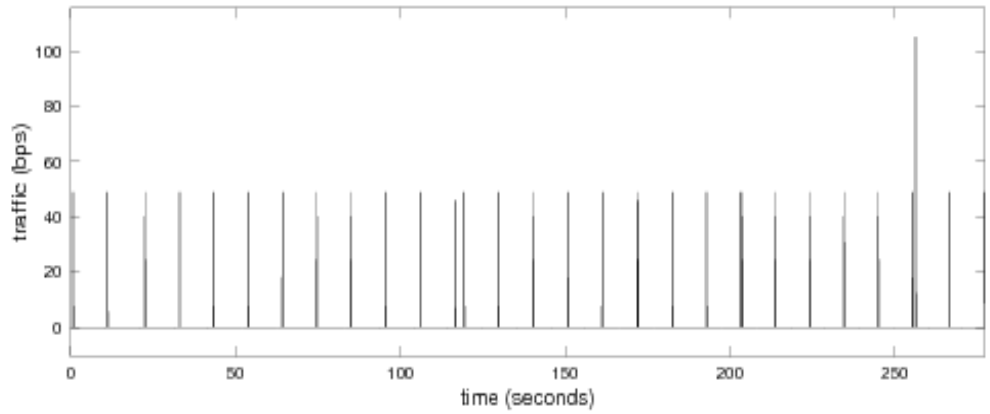

Figura 4.16 - Tráfego *downstream* POP3 por parte do cliente na direção B (bytes por segundo).

Analisando a Figura 4.18, observa-se a existência de um pico de tráfego por volta dos três minutos, o que pode significar a resposta do utilizador a acusar a receção de um email. O restante tráfego está agregado em intervalos de tempo extremamente pequenos (inferiores a dez segundos) e é maioritariamente tráfego de controlo.

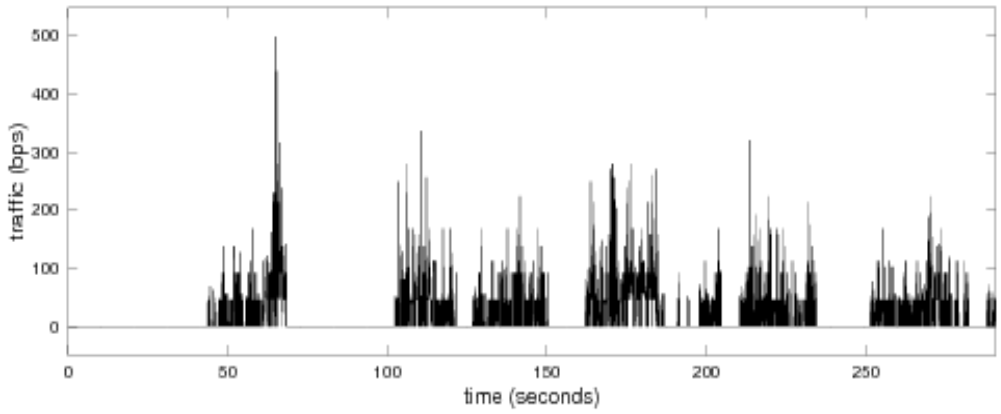

Figura 4.17 - Tráfego *upstream* POP3 por parte do cliente na direção A (bytes por segundo)

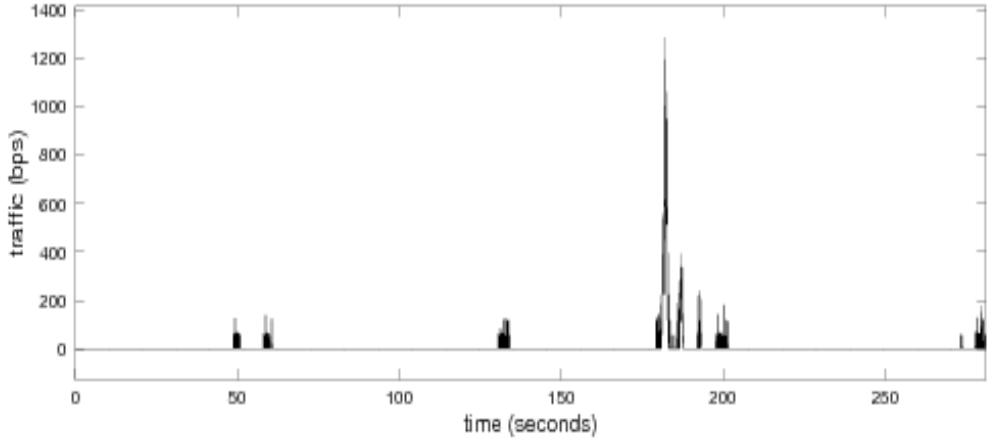

Figura 4.18 - Tráfego *upstream* POP3 por parte do cliente na direção B (bytes por segundo).

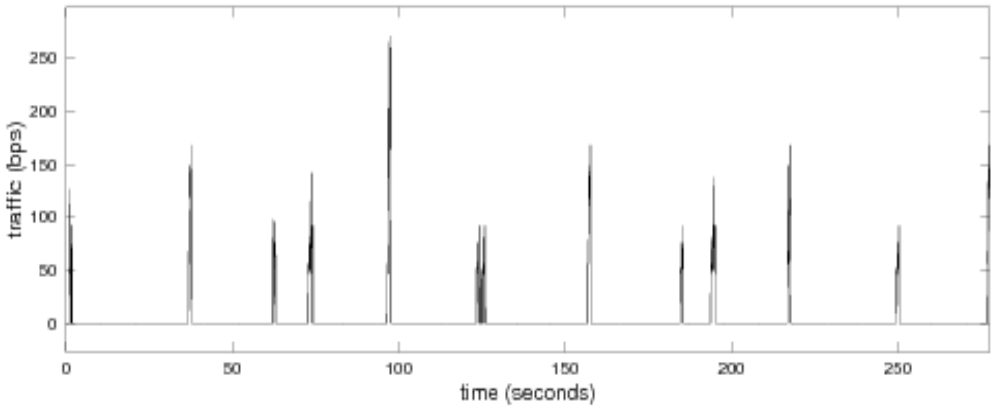

Figura 4.19 - Tráfego *upstream* POP3 por parte do cliente na direção B (bytes por segundo).

Analisando a Figura 4.19, verifica-se a existência de tráfego de pacotes de pequenas dimensões durante breves instantes e separados por intervalos de tempo irregulares. Tendo em conta as dimensões dos pacotes capturados, este tráfego será maioritariamente de controlo.

4.1.4 IMAP

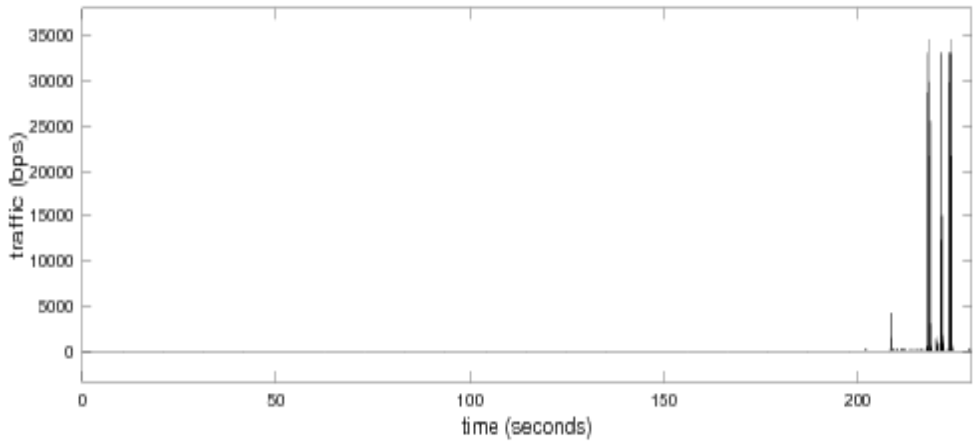

Figura 4.20 - Tráfego *downstream* IMAP por parte do cliente na direção A (bytes por segundo).

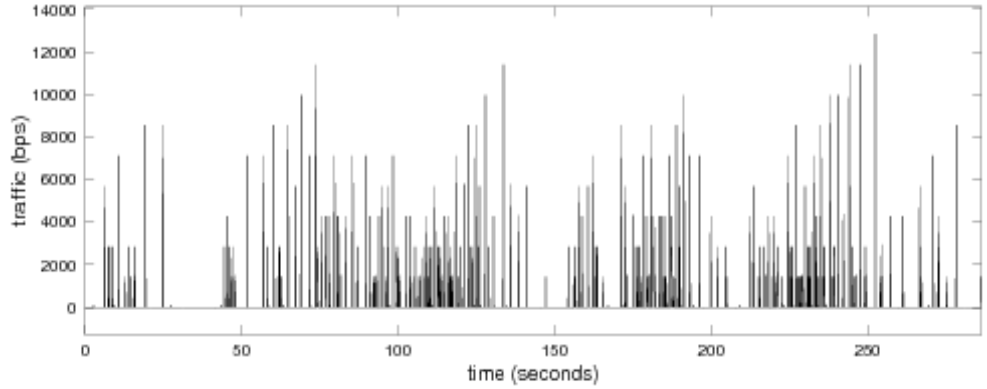

Figura 4.21 - Tráfego *downstream* IMAP por parte do cliente na direção B (bytes por segundo).

Juntamente com o POP3, o IMAP (*Internet Message Access Protocol*) é o protocolo mais usado para a recepção de emails. É um protocolo da camada de Aplicação que utiliza o porto TCP 143. O IMAP trabalha tanto online como *offline*.

As mensagens de email recebidas pelo utilizador são encaminhadas para um servidor de emails que os guarda na caixa de correio do utilizador. O utilizador pode aceder a estas mensagens usando uma aplicação cliente de email. Atualmente, a maior parte dos clientes de email têm suporte para POP3 e IMAP.[51]

Estudam-se agora alguns exemplos de tráfego IMAP existente nas capturas de tráfego em investigação. A Figura 4.20 tem características normalmente associadas ao tráfego gerado pelos utilizadores quando acedem à sua caixa de email: a existência de picos de tráfego bastante próximos indicia a receção de um ou mais emails de dimensão elevada (possivelmente com anexos, dada a dimensão dos pacotes).

É possível observar-se na Figura 4.21 a ocorrência de vários picos de tráfego, sempre acompanhados de tráfego contínuo de pacotes de grande dimensão, existindo poucos períodos sem fluxo de pacotes *downstream*. É possível assumir-se que o utilizador efetuou o download de alguns emails, dado o elevado volume de tráfego.

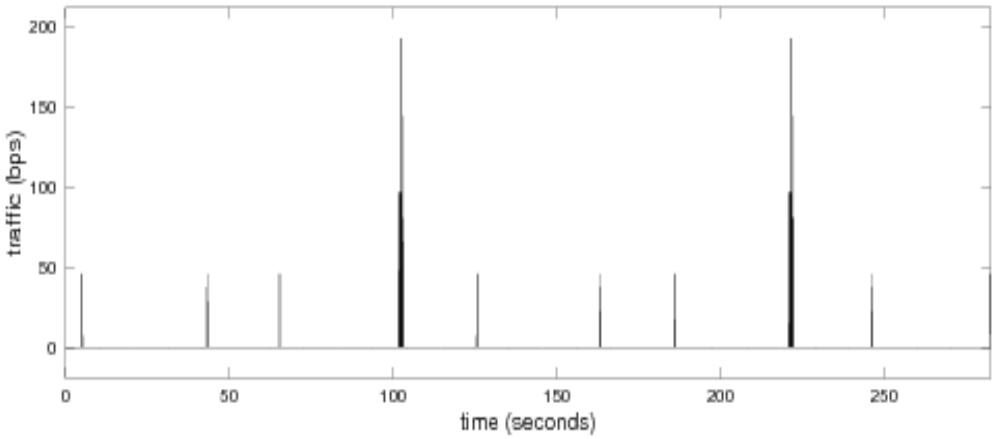

Figura 4.22 - Tráfego *downstream* IMAP por parte do cliente na direção B (bytes por segundo).

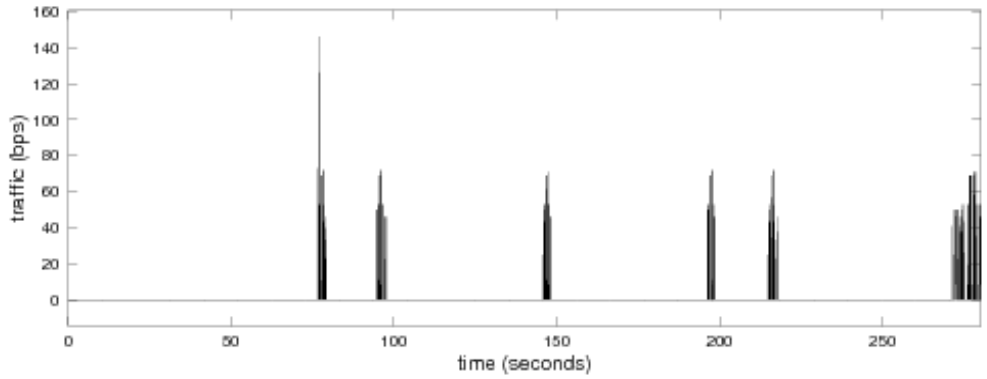

Figura 4.23 - Tráfego *upstream* IMAP por parte do cliente na direção A (bytes por segundo).

Observando a Figura 4.22, verifica-se um fluxo reduzido de pacotes de pequenas dimensões, o que significa a inexistência da chegada de novos emails. Este tráfego será predominantemente tráfego de controlo.

No que diz respeito ao tráfego *upstream* IMAP, foi possível observar dois casos distintos. Na Figura 4.23 o tráfego de pacotes é reduzido, sendo que é possível reconhecer a existência de grupos de pacotes que surgem de forma sequencial durante alguns segundos. Estas sequências surgem separadas por intervalos de tempo elevados. Tendo em conta as dimensões dos pacotes observados, este tráfego não será de dados mas principalmente de controlo das sessões IMAP abertas.

Relativamente à Figura 4.24, é possível observar três momentos distintos de fluxo de tráfego; cada uma destas sequências tem durações diferentes e alguns picos de tráfego, o que indicia que este tráfego pode corresponder à confirmação da consulta da caixa de correio por parte do utilizador.

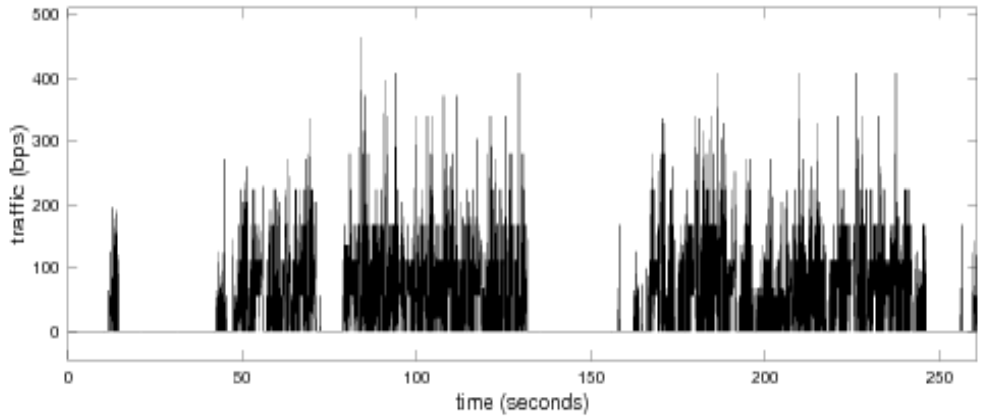

Figura 4.24 - Tráfego *upstream* IMAP por parte do cliente na direção B (bytes por segundo).

4.1.5 RTSP

O RTSP (*Real Time Streaming Protocol*) é um protocolo da camada de Aplicação de controlo de conteúdos multimédia em tempo real. Este protocolo utiliza o porto 554.Tanto é capaz de processar ficheiros guardados como transmitir conteúdos em direto e foi dimensionado para controlar sessões múltiplas de transmissão de dados. O RTSP garante os meios para que a transmissão possa ser efetuada por UDP (embora nas capturas de tráfego utilizadas este tipo de transmissão não esteja presente) ou TCP, *multicast* ou *unicast*.

A transmissão dos dados em si não é assegurada pelo protocolo RTSP, mas sim pela colaboração com os protocolos *Real Time Transport Protocol* (RTP) e *Real Time Control Protocol* (RTCP). [52, 53]

A sessão RTSP é iniciada pelo cliente estabelecendo uma ligação TCP com o servidor, geralmente direcionada para o porto 554. De seguida, o cliente interage com o servidor enviando uma série de comandos para planear a transmissão de dados. O RTSP é semelhante ao HTTP em certos aspetos, pois existem comandos RSTP parecidos com os *HTTP Request* que permitem ao utilizador controlar os dados multimédia que recebe: *OPTIONS* (permite saber os pedidos que o servidor aceita); *DESCRIBE* (inclui a descrição dos dados que podem ser transferidos); SETUP (este pedido deve ser sempre enviado antes da opção PLAY. Especifica como o *stream* deve ser enviado e os portos locais de receção dos dados); PLAY (ordem para a reprodução de um ou mais *streams*, dependendo se o endereço pretendido for só um ou se for um agregado); PAUSE (permite interromper temporariamente a transmissão, que é reatada com um pedido PLAY. O pedido inclui o endereço do *stream* em causa e pode incluir um parâmetro que especifique quando deve ser efetuada a pausa na transmissão). Quando o cliente pretende finalizar a transmissão, envia o comando TEARDOWN assim como o ID da sessão para que o servidor cancele a transmissão de dados. [52, 53]

De seguida apresentam-se alguns exemplos de tráfego RTSP existente nas capturas de tráfego em estudo. Na Figura 4.25 verifica-se tráfego constante e contínuo, sem picos de tráfego, normalmente associado à visualização de vídeos online. [44]

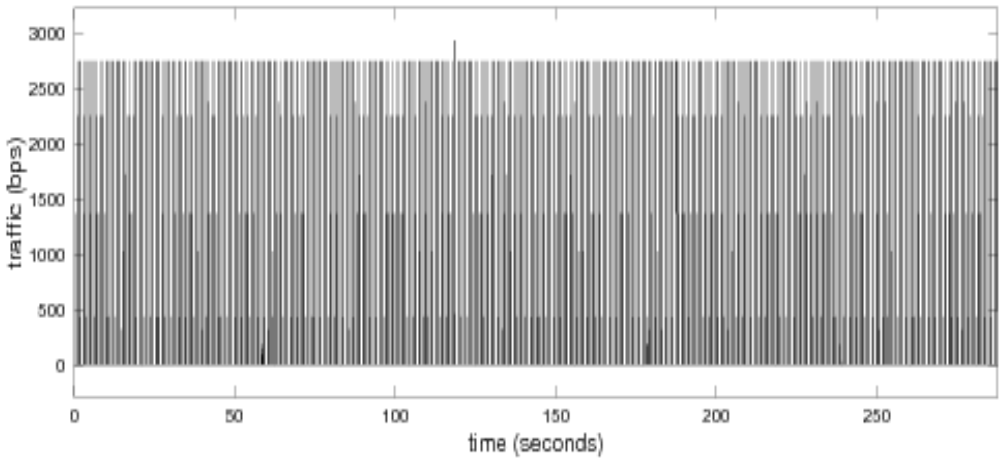

Figura 4.25 - Tráfego *downstream* RTSP por parte do cliente na direção B (bytes por segundo).

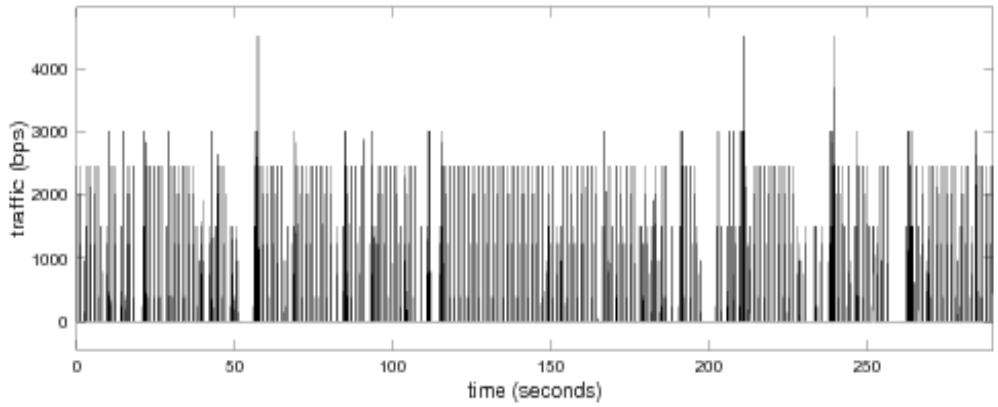

Figura 4.26 - Tráfego *downstream* RTSP por parte do cliente na direção A (bytes por segundo).

No que diz respeito à Figura 4.26, verifica-se que o fluxo de tráfego é particularmente constante durante a maior parte do tempo da amostra, havendo exceções que são os períodos de tempo sem qualquer tráfego, os quais são sucedidos por picos de tráfego. Os poucos picos de tráfego existentes na amostra iniciam novas sequências de tráfego *downstream*.

No que concerne ao tráfego *upstream* RTSP foi possível observar dois casos distintos. No caso da Figura 4.27, estamos na presença de tráfego contínuo de pacotes de dimensões semelhantes (mas menores comparativamente aos exemplos do tráfego *downstream* RTSP) e sem picos de tráfego.

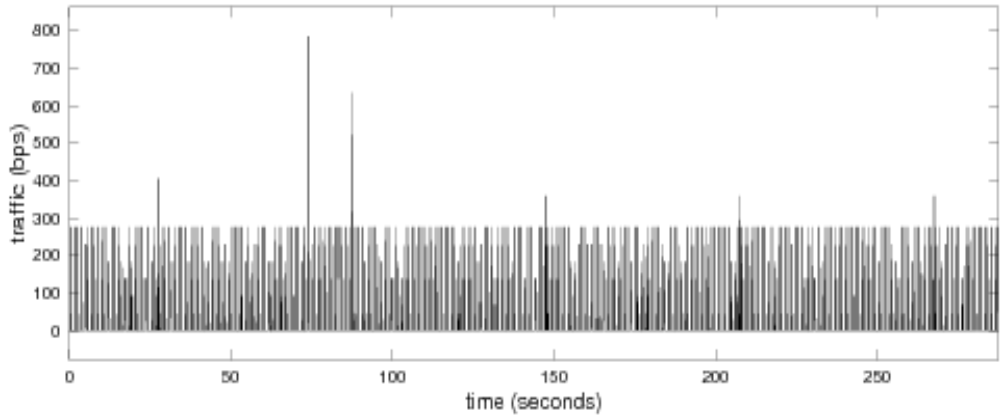

Figura 4.27 - Tráfego *upstream* RTSP por parte do cliente na direção B (bytes por segundo).

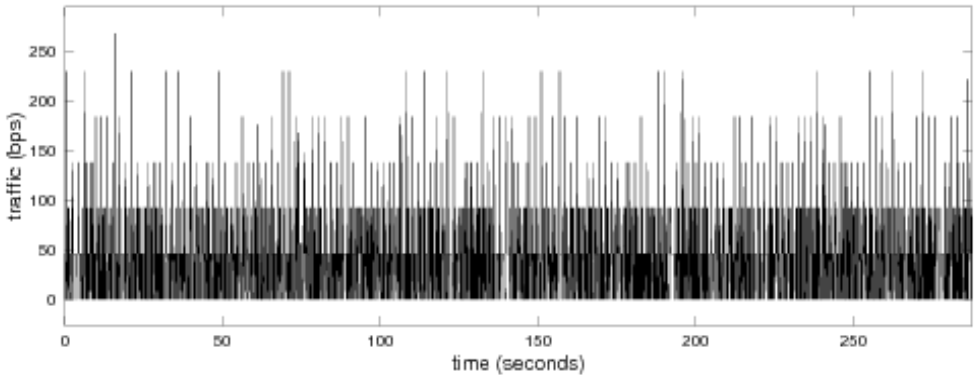

Figura 4.28 - Tráfego *upstream* RTSP por parte do cliente na direção B (bytes por segundo).

Relativamente à Figura 4.28, é possível observar um grande volume de tráfego de forma contínua, registando-se a ocorrência ocasional de um pico de tráfego logo nos primeiros instantes da amostra. Não existem falhas ao longo da janela temporal da figura.

4.1.6 MSNP

O protocolo MSNP (*Mobile Status Notification Protocol* ou *Microsoft Notification Protocol*) é um protocolo de *instant messaging* da camada de aplicação desenvolvido pela Microsoft para ser usado pelo *Microsoft Messenger Service* e aplicações de *instant messaging* ligadas ao mesmo como o *Windows Live Messenger*. À medida que a Microsoft foi lançado novas edições do *Windows Live Messenger*, o protocolo MSNP acompanhou essa evolução, sendo a última versão o MSNP 19. Nas últimas versões, o MSNP já tem capacidade para *streaming* de vídeo, chamadas e VoIP. [54, 55]

Certos clientes de *instant messaging* fora da alçada da Microsoft também conseguem comunicar através deste protocolo, como por exemplo o Pidgin. O MSNP utiliza o porto 1863, tanto sobre TCP como sobre UDP (embora nas capturas de tráfego utilizadas nesta dissertação, o MSNP operava apenas sobre UDP).

As sessões deste protocolo consistem numa sequência de comandos trocados entre o cliente de *instant messaging* e o servidor. Sempre que houver alguma atividade dos contactos do utilizador (contacto entra online, contacto inicia conversação, etc), o servidor envia um comando ao cliente com um aviso da ocorrência, de modo a que esta informação apareça na interface do utilizador. [54, 55]

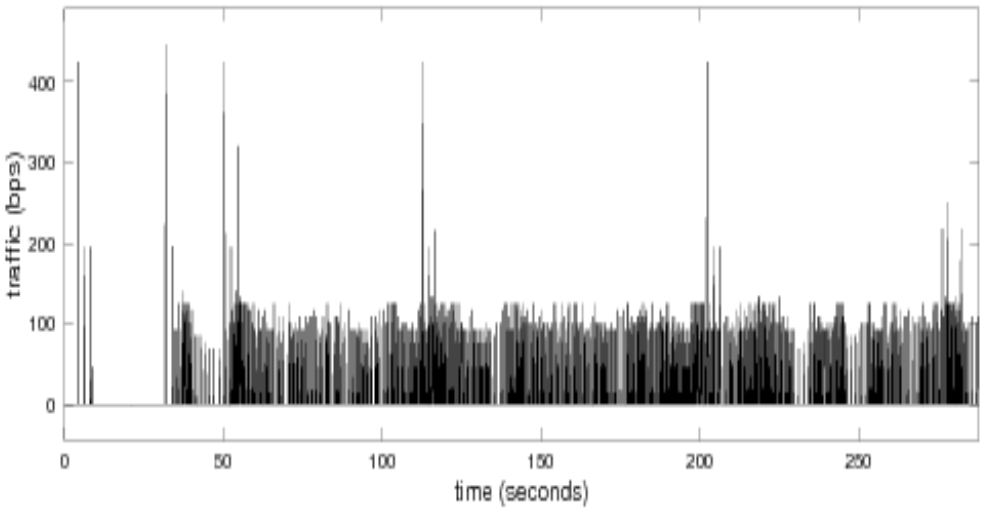

Figura 4.29 - Tráfego *downstream* MSNP por parte do cliente na direção A (bytes por segundo).

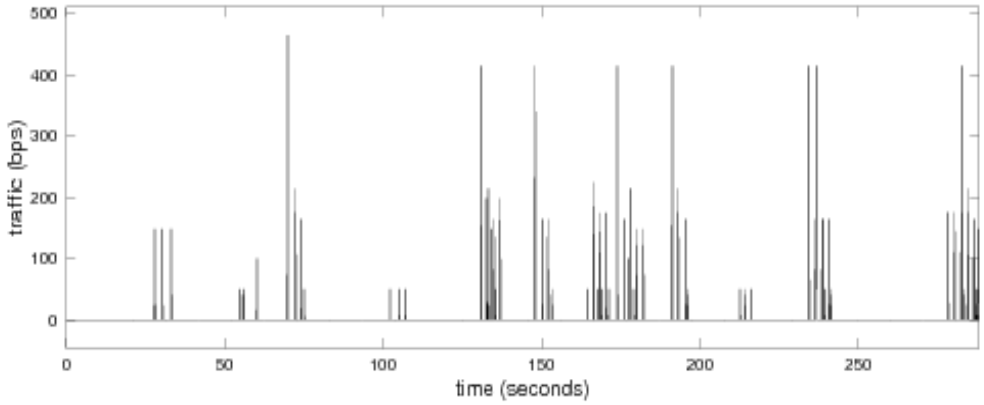

Figura 4.30 - Tráfego *downstream* MSNP por parte do cliente na direção A (bytes por segundo).

De seguida apresentam-se os exemplos de tráfego MSNP presentes nas capturas de tráfego analisadas. No que concerne à Figura 4.29, verifica-se a existência de um pico de tráfego nos primeiros instantes da amostra. De seguida, temos alguns instantes sem qualquer tráfego, até que surgem novos picos de tráfego e a partir daí o fluxo de pacotes torna-se contínuo durante o restante intervalo da janela temporal considerada.

Em relação à Figura 4.30, verifica-se a ocorrência de vários picos de tráfego ao longo do tempo, após os quais surge fluxo durante alguns segundos.

Relativamente ao tráfego *upstream* MSNP, encontram-se dois exemplos com características discrepantes. O primeiro exemplo (Figura 4.31) apresenta tráfego contínuo ao longo da janela temporal da amostra, à exceção de instantes sem qualquer fluxo que ocorrem aproximadamente uma vez por minuto. Regista-se a ocorrência de alguns picos de tráfego ao longo do tempo.

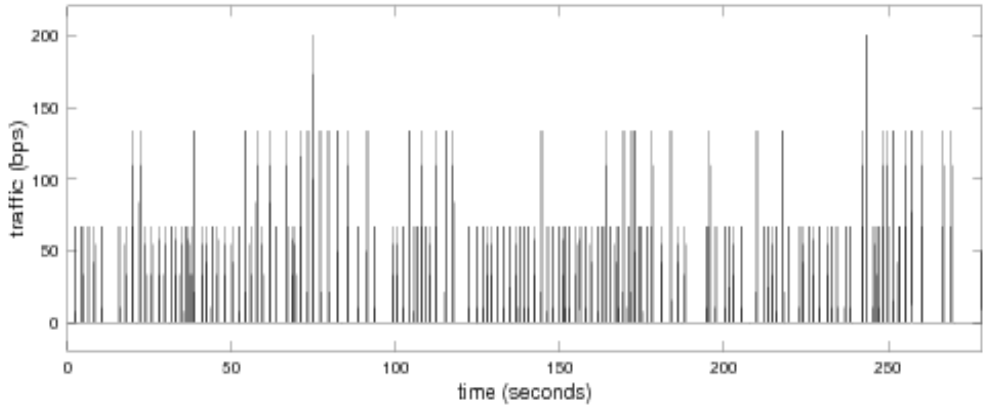

Figura 4.31 - Tráfego *upstream* MSNP por parte do cliente na direção A (bytes por segundo).

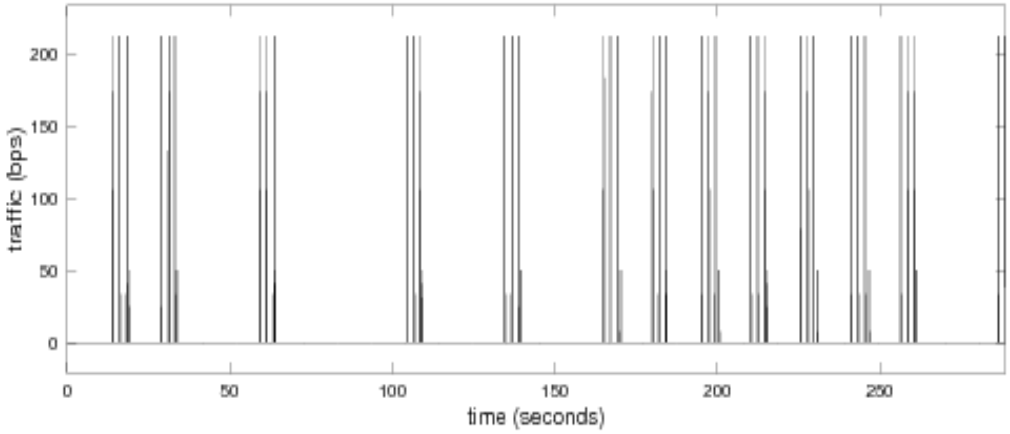

Figura 4.32 - Tráfego *upstream* MSNP por parte do cliente na direção A (bytes por segundo).

O segundo exemplo (Figura 4.32) apresenta picos de tráfego periódicos com duração e características semelhantes, separados entre si de forma irregular. Estes picos de tráfego desvanecem-se de forma rápida, pois duram apenas alguns segundos.

4.1.7 XBOX

O serviço *XBOX Live* é uma plataforma de jogos online e outros serviços multimédia lançada pela Microsoft em 2002 com o propósito inicial de possibilitar que os jogadores das consolas da empresa (primeiro a XBOX e depois da sua descontinuação, a XBOX 360) pudessem jogar online em modo *multiplayer* os jogos desenvolvidos para essas consolas. É possível aceder a esta plataforma através das consolas ou com um computador pessoal e está disponível em trinta e sete países. A biblioteca de jogos disponibilizada atualmente pelo *XBOX Live* é superior a mil jogos, de diferentes géneros: aventura, desporto, estratégia, música, plataformas, RPG, FPS, etc. [56]

Ao longo dos anos, a Microsoft adicionou novas funcionalidades a este serviço, embora algumas delas só estejam disponíveis mediante subscrição: serviço de Messenger (MSN Messenger), motor de busca (Bing), browser (Internet Explorer), acesso a redes sociais (Skype, Facebook, Twitter, last.fm, etc.), disponibilização de conteúdos *on-demand* (filmes, vídeos e séries) e *streaming* de conteúdos (*XBOX Music*, *XBOX*

Video e Netflix). [56]

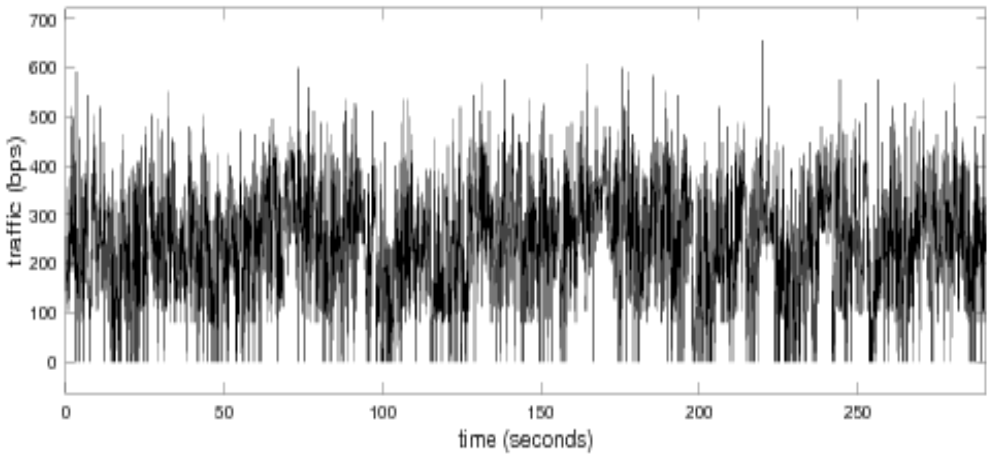

Figura 4.33 - Tráfego *downstream* XBOX por parte do cliente na direção A (bytes por segundo).

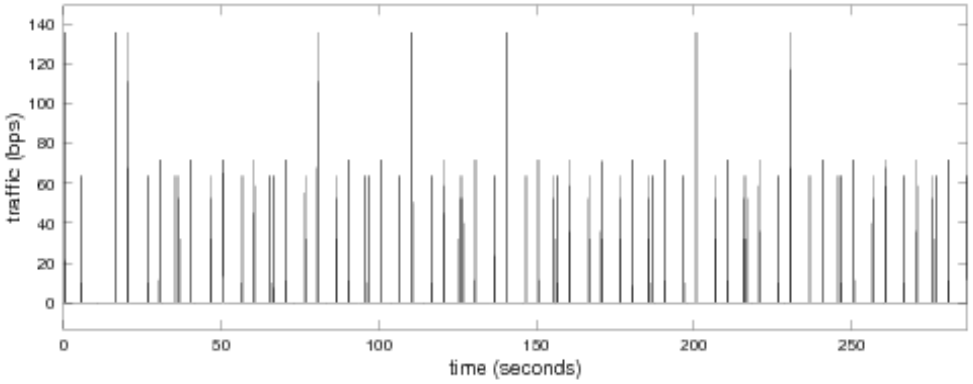

Figura 4.34 - Tráfego *downstream* XBOX por parte do cliente na direção A (bytes por segundo).

Este protocolo tem atribuído o porto 3074, tanto para TCP como para UDP (embora nas capturas de tráfego usadas no âmbito desta dissertação apenas foi capturado tráfego a operar sobre UDP). [56]

Apresentam-se agora alguns exemplos de tráfego XBOX existente nas capturas em investigação. No caso do tráfego *downstream*, o primeiro exemplo (Figura 4.33) apresenta tráfego irregular mas contínuo (característico da visualização de vídeos e animações interativas) e sem qualquer paragem.

O segundo exemplo (Figura 4.34) apresenta alguns picos de tráfego mas também fluxo periódico de pacotes de tamanho reduzido, o que é um cenário bastante distinto do primeiro exemplo.

No terceiro exemplo (Figura 4.35) encontra-se tráfego constante ao longo de toda a janela temporal, pontuado com vários picos de tráfego, principalmente no primeiro minuto e no último minuto.

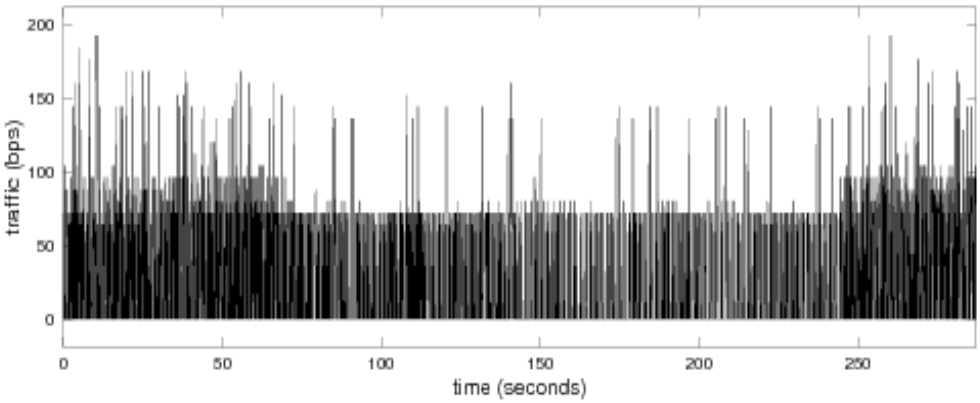

Figura 4.35 - Tráfego *downstream* XBOX por parte do cliente na direção B (bytes por segundo).

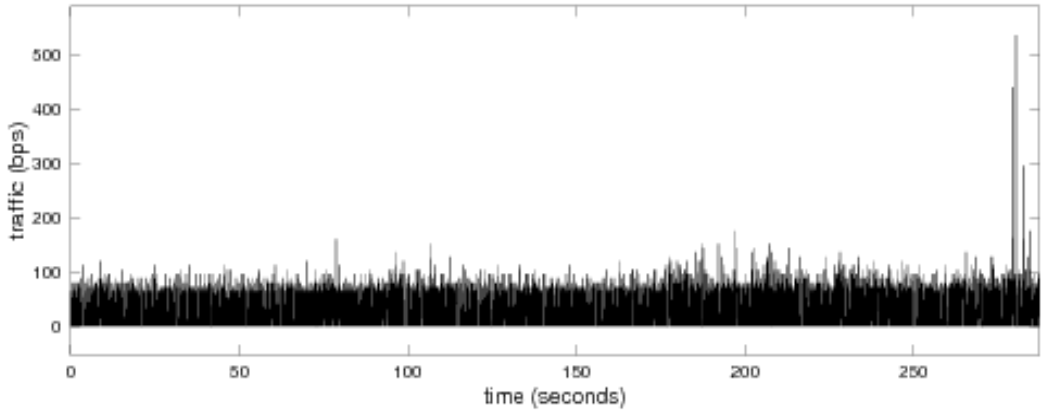

Figura 4.36 - Tráfego *upstream* XBOX por parte do cliente na direção B (bytes por segundo).

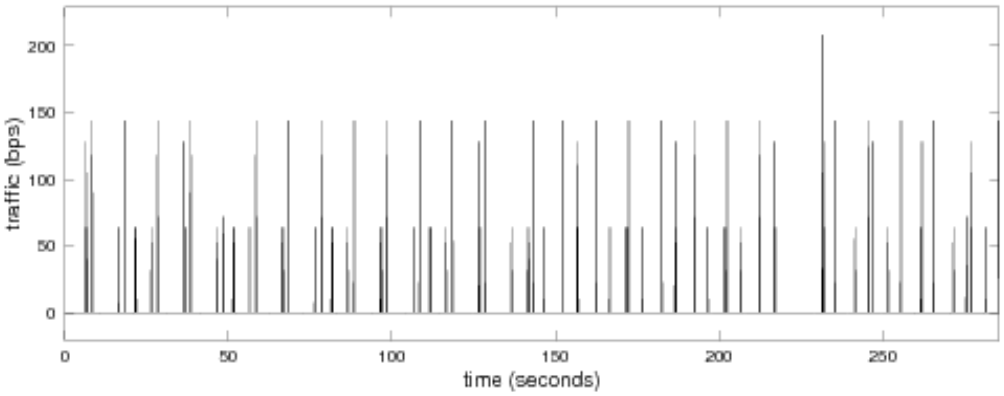

Figura 4.37 - Tráfego *upstream* XBOX por parte do cliente na direção B (bytes por segundo).

No que concerne ao tráfego *upstream*, surgiram dois exemplos relevantes. No primeiro caso (Figura 4.36) temos tráfego contínuo e de características semelhantes durante todo o intervalo temporal desta figura. Sensivelmente nos últimos dez segundos surgiram picos de tráfego.

No segundo exemplo (Figura 4.37) verifica-se a ocorrência de numerosos picos de tráfego ao longo do intervalo temporal da amostra. O tráfego é irregular, pois os picos de tráfego são intercalados por períodos de tempo sem qualquer atividade.

4.2 Processamento de Tráfego

O processamento das capturas de tráfego requeridos para estudo neste trabalho começou pela adaptação dos ficheiros ".pcap" para que o seu tratamento fosse mais fácil. Tendo em conta que as capturas de tráfego de cada direção estavam divididas em cinco ficheiros ".pcap", foi necessário uni-las de forma a obter-se um único ficheiro ".pcap" para cada direção, recorrendo ao comando "mergecap ficheiro1.pcap ficheiro2.pcap ficheiro3.pcap ficheiro4.pcap ficheiro5.pcap -w ficheiro_saída.pcap". Como estes ficheiros são muito pesados para tratamento no *Wireshark*, procedeu-se à divisão de ambos as capturas de tráfego em ficheiros ".pcap" mais pequenos. Em cada direção, cada protocolo tratado neste trabalho (HTTP, SMTP, IMAP, POP3, RTSP, MSNP, XBOX) terá dois ficheiros ".pcap"; um ficheiro ".pcap" em que o porto de destino do tráfego é o porto do protocolo respetivo e outro ficheiro ".pcap" em que o porto de origem do tráfego é o porto do protocolo. Exemplificando o caso do HTTP (porto de serviço é o 80), obtém-se os ficheiros ".pcap" com os comandos "tshark -r ficheiro_input.pcap "tcp.dstport==80" -w ficheiro_output_dst80.pcap" e "tshark -r ficheiro_input.pcap "tcp.srcport==80" -w ficheiro_output_src80.pcap". Assim é possível dividir a análise dos resultados em quatro blocos: Upload de dados por parte do cliente, Download de dados por parte do servidor, Download de dados por parte do cliente e Upload de dados por parte do servidor.

Uma vez que um dos objetivos desta dissertação é diferenciar fluxos de dados e associá-los a diferentes aplicações, tendo em conta os seus padrões de transmissão, o processamento dos ficheiros ".pcap" dos vários protocolos vai ser efetuado recorrendo à ferramenta *Tshark*. Para cada um dos quatro blocos, foram escolhidos aleatoriamente uma série de fluxos tendo em conta a duração da atividade dos utilizadores ao longo do tempo. Cada fluxo será identificado pelo endereço IP da fonte ou do destino, dependendo do bloco em causa; assim, será necessário um comando diferente para cada bloco.

Foi então criado um programa em *Python*, onde se incluem os comandos *tshark* necessários para a obtenção das estatísticas pretendidas e ainda formatação dos ficheiros de saída. Usando endereços IP exemplo para cada bloco e considerando o caso do protocolo HTTP, os comandos para o programa *Tshark* usados foram os seguintes:

- Cliente (*Upstream*) - "tshark -r ficheiro.pcap -qz io,stat,0,1, ip.src==1.1.1.1&&tcp.dstport==80"

- Servidor (*Downstream*) - "tshark -r ficheiro.pcap -qz io,stat,0,1, ip.dst==1.1.1.1&&tcp.dstport==80"

- Cliente (*Downstream*) - "tshark -r ficheiro.pcap -qz io,stat,0,1, ip.dst==1.1.1.1&& tcp.srcport==80"

- Servidor (*Upstream*) - "tshark -r ficheiro.pcap -qz io,stat,0,1, ip.src==1.1.1.1&& tcp.dstport==80"

Para cada um dos outros protocolos, procedeu-se de forma análoga. A aplicação destes comandos permitiu obter dados estatísticos sobre o volume total de pacotes e bytes em função do tempo a cada 100 milissegundos. O programa criado permitia a edição dos ficheiros de dados, retirando o seu cabeçalho, restando apenas as colunas com os dados importantes para os resultados finais (o método de obtenção dos

resultados apresentados na Análise e Discussão dos Resultados é explicado de seguida no subcapítulo 4.3).

4.3 Análise de Tráfego

Após o processamento dos dados, é necessário proceder-se à análise dos mesmos; para o efeito construíram-se gráficos de modo a facilitar a compreensão dos resultados obtidos. Foram escolhidos os escalogramas e a diferenciação multi-escalar dos fluxos de dados para a apresentação dos resultados (a componente teórica de ambos já foi abordada no capítulo 3). Apesar dos diferentes intervalos de tempo utilizados consoante o tipo de tráfego, estabeleceu-se a utilização de Bps (bytes por segundo) em todos os escalogramas.

Para a obtenção do escalograma de cada fluxo de dados, foi criado um programa em *Octave* que recorrendo a bibliotecas de *wavelets*, cria escalogramas com escala em frequência 1:128.

Para a obtenção da diferenciação multi-escalar de cada fluxo de dados e posterior comparação com outros fluxos foram criados dois programas em *Octave*: o primeiro programa calcula o desvio padrão normalizado da energia ao longo do tempo para cada fluxo; o segundo programa agrupa vários fluxos na mesma figura, diferenciando-os entre si usando cores e traçados diferentes.

5 Análise e Discussão dos Resultados

A análise dos resultados é uma tarefa complexa devido à dimensão do cenário e da quantidade de parâmetros em estudo. O estudo realizado neste trabalho focou-se em tráfego de duas capturas de tráfego, com fluxo de pacotes em direções opostas; foram estudados quatro contextos – tráfego *upstream* por parte do cliente, tráfego *downstream* por parte do servidor, tráfego *downstream* por parte do cliente e tráfego *upstream* por parte do servidor – de sete protocolos diferentes (HTTP, SMTP, POP3, IMAP, RTSP, MSNP e XBOX). Assim, tendo em conta a quantidade de resultados obtidos cada protocolo vai ser analisado individualmente, efetuando-se posteriormente a comparação entre as suas características.

Em cada protocolo apresentam-se as métricas do tráfego capturado, amostrado em intervalos de 0,1 segundos (com a taxa de download /upload de tráfego em Bps) e com os respetivos escalogramas. Cada uma destas figuras apresenta características únicas relacionadas com padrões de tráfego distintos provocados pela interação humana com os serviços de rede. Apresentam-se também os gráficos do desvio padrão da análise da diferenciação multi-escalar ao longo do tempo, tendo em conta a respetiva escala de frequência. Através destas figuras, analisando a forma como a energia do processo varia ao longo da gama de frequências é possível diferenciar cada fluxo de tráfego, agrupando-os segundo o seu comportamento. Tendo em conta a anonamização das capturas de tráfego em estudo, a tarefa de associar cada fluxo à aplicação original que o originou é mais complicada.

5.1 HTTP

5.1.1 Cliente (*Downstream*)

Para o download de tráfego HTTP por parte do cliente, encontraram-se cinco casos distintos. No primeiro caso (Figura 5.1), ocorrem alguns picos não periódicos de curta duração e grande amplitude, mas apenas no espaço de cerca de um minuto. Estes picos são causados pelos cliques dos utilizadores quando selecionam uma nova página para visualização, gerando também um número apreciável de componentes de baixa frequência. Surgem também alguns componentes de média frequência, devido à criação de sessões TCP, e componentes de alta frequência devido à chegada de pacotes.

No caso seguinte (Figura 5.2), regista-se a ocorrência de picos de grande amplitude de uma forma mais regular comparativamente à Figura 5.1, significando uma atividade mais intensa do utilizador. Verifica-se que não existem muitos componentes de média e alta frequência, o que leva a crer que esta figura pode representar um exemplo de tráfego gerado por aplicações de redes sociais. Nestas situações, os picos de tráfego normalmente são associados a atualizações de estados de outros utilizadores das redes sociais ligados online que surgem nos *feeds* de notícias ou também à visualização dos perfis de outros utilizadores.

No que concerne à Figura 5.3, verifica-se a existência de vários picos de tráfego pseudo periódicos, sendo grande parte deles de baixa amplitude. Este tipo de tráfego normalmente surge associado à visualização de fotografias online, em que a periodicidade dos picos de tráfego está relacionada com os cliques do utilizador quando acede a uma nova fotografia. A presença de muitos componentes de baixa frequência no escalograma advém dos repetidos cliques do utilizador. Contudo, este comportamento do tráfego também é característico do download de vídeos longos no Youtube, sendo que

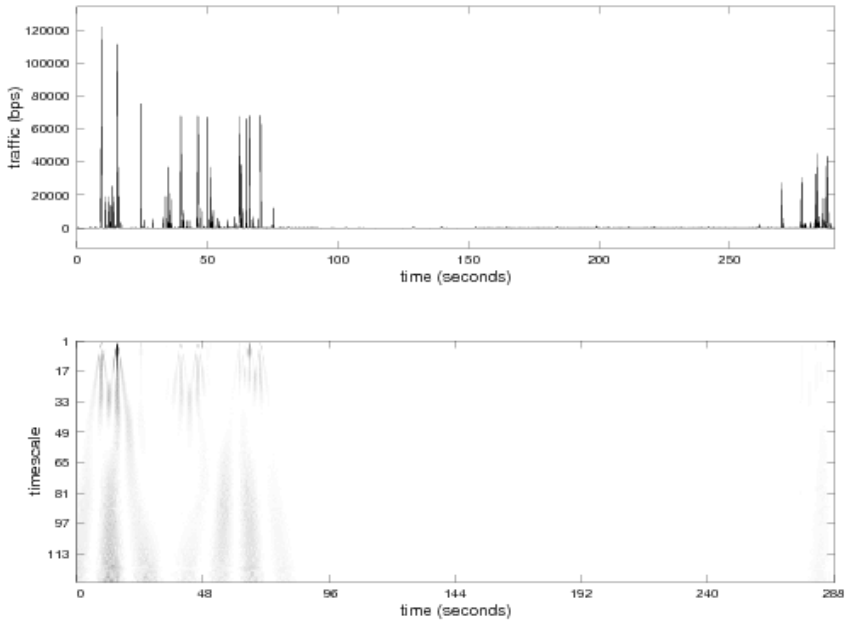

Figura 5.1 – Tráfego downstream HTTP por parte do cliente na direção B (bytes por segundo).

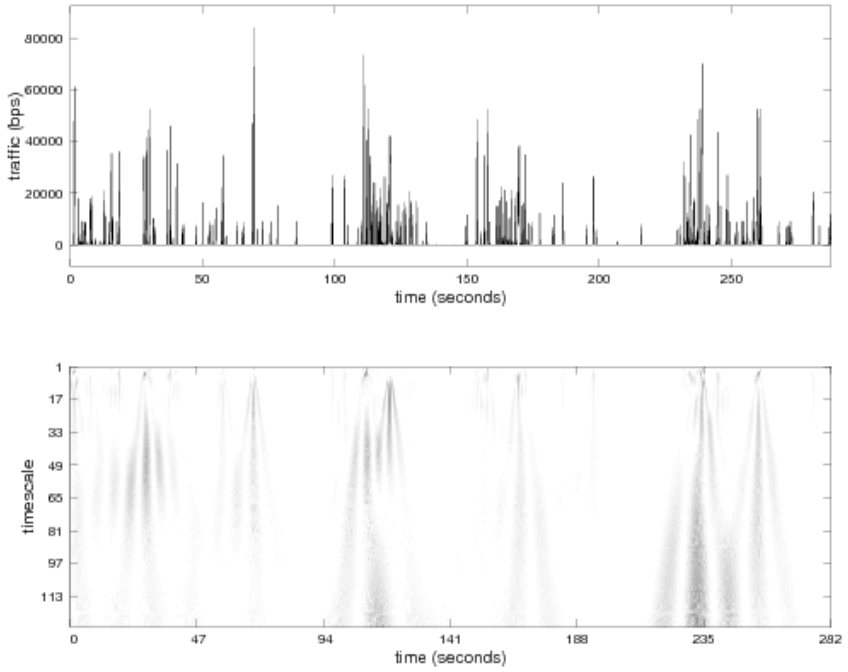

Figura 5.2 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

cada pico pseudo periódico corresponde ao pedido do enchimento do buffer por parte do cliente. O buffer do vídeo nesta aplicação depende do seu tamanho, pois inicialmente é efetuado um carregamento considerável de dados (de modo a iniciar a transmissão do vídeo) e de seguida é enviado um fluxo de dados contínuo, preenchendo o buffer e impedindo falhas na transmissão do vídeo.

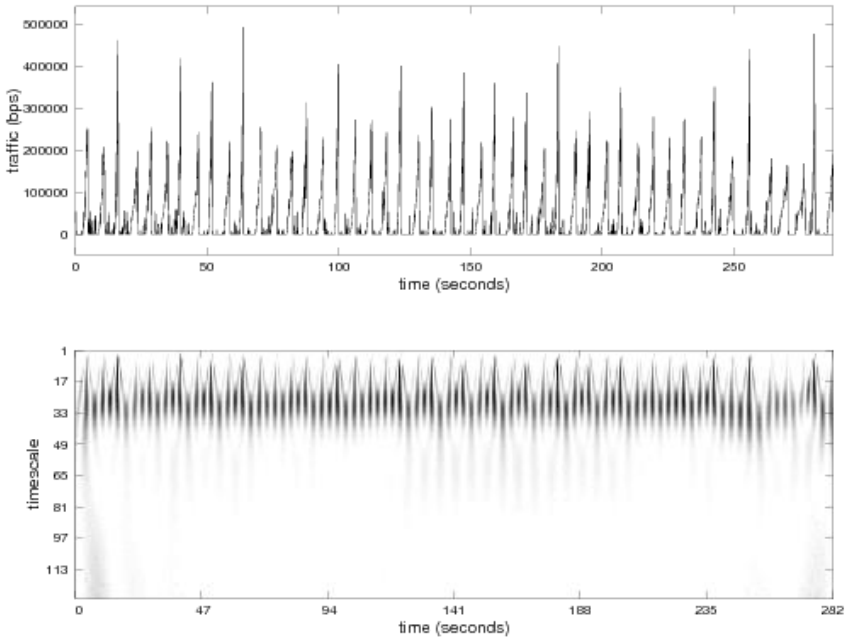

Figura 5.3 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

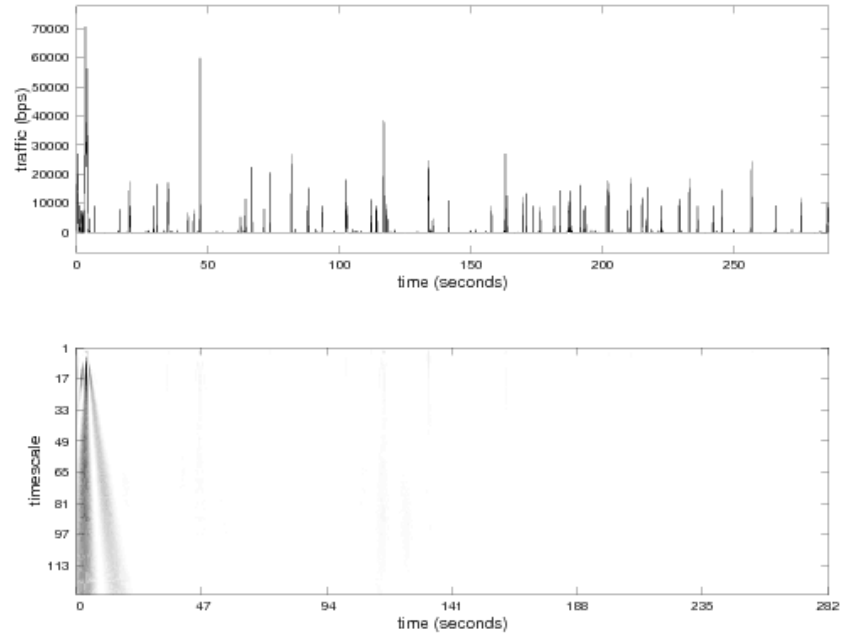

Figura 5.4 - Tráfego *downstream* HTTP por parte do cliente na direção B (bytes por segundo).

A Figura 5.4 apresenta picos de tráfego pouco frequentes, não periódicos e de baixa amplitude. Os componentes de frequência visíveis no escalograma surgem no início da escala temporal, associados a um pico de tráfego de grande amplitude. A presença destes componentes desta forma indicia que o utilizador ao efetuar *browsing* solicitou a abertura de sessões TCP, o que originou consequentemente a chegada de pacotes, como resposta ao seu pedido.

Finalmente, a Figura 5.5 representa tráfego normalmente associado à partilha de ficheiros (utilizando neste caso o porto 80) ou ao download de ficheiros de grande

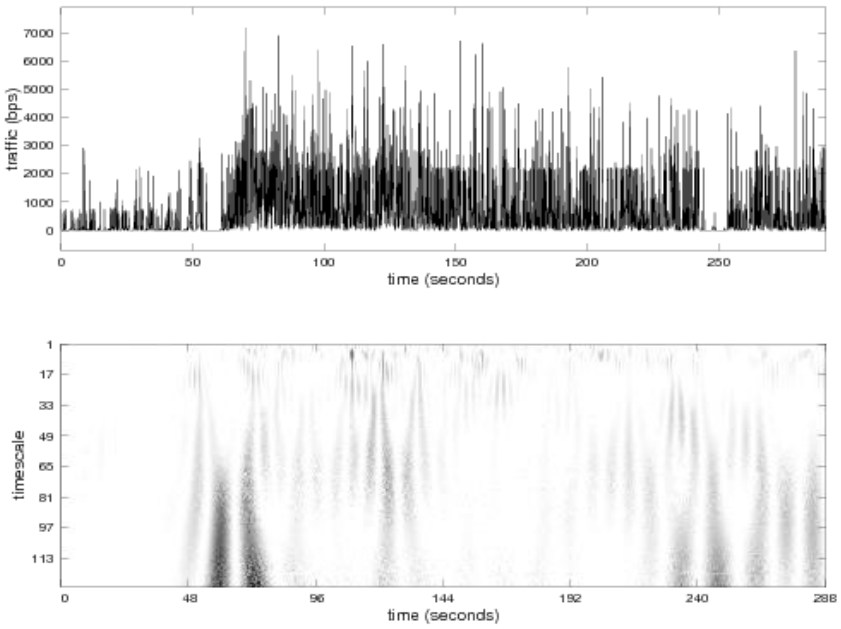

Figura 5.5 - Tráfego *downstream* HTTP por parte do cliente na direção A (bytes por segundo).

tamanho: tráfego de grande largura de banda com pouco tempo a mediar a chegada de cada pacote (*inter-arrival time* baixo), devido ao download contínuo do conteúdo. Existem alguns picos de tráfego quando a largura de banda é maximizada para o download do conteúdo pretendido e muitos componentes de alta frequência, devido à constante chegada de pacotes. Regista-se a pouca presença de componentes de baixa frequência, pois os cliques de utilizador não são relevantes durante a transferência deste tipo de conteúdos.

Analisando a Figura 5.6 cada região demarcada caracteriza um certo segmento de frequências contendo eventos com uma gama de variação de energia bastante específica. Cada uma destas regiões está associada a determinados eventos gerados por atividade humana ou pela rede. A região A, por exemplo, engloba eventos de baixa frequência com uma variação de energia moderada. Estes eventos são normalmente gerados por cliques de utilizadores ao tentarem aceder a novas páginas em sites de notícias online, *browsing* de fotografias em sites de partilha de fotografias e as interações que ocorrem em aplicações de redes sociais (Fluxos 2,3,11,14 e 16). A região B, por sua vez, abrange apenas eventos de baixa frequência com uma variação de energia pequena, como são exemplo os sites de visualização de vídeo online. A região C contém dois fluxos (8 e 11) e abrange eventos de média frequência com elevada variação de energia, devido à criação em grande número de sessões TCP e HTTP. Significa então que estes dois fluxos estão associados a uma atividade de *browsing* intensa por parte dos utilizadores em causa. A região D envolve eventos de média frequência com variação de energia menor comparativamente à região C e contém uma percentagem maioritária dos fluxos considerados neste cenário. Já a região E, apesar de também se situar no segmento das médias frequências, compreende eventos com uma variação de energia muito reduzida, ou seja, existem poucas interações ao nível de sessões TCP e HTTP, o que é normalmente característico das aplicações de redes sociais (Fluxos 10 e 14).

Relativamente ao segmento das altas frequências, é possível delimitar duas regiões. A região F engloba eventos com variação de energia pequena a moderada.

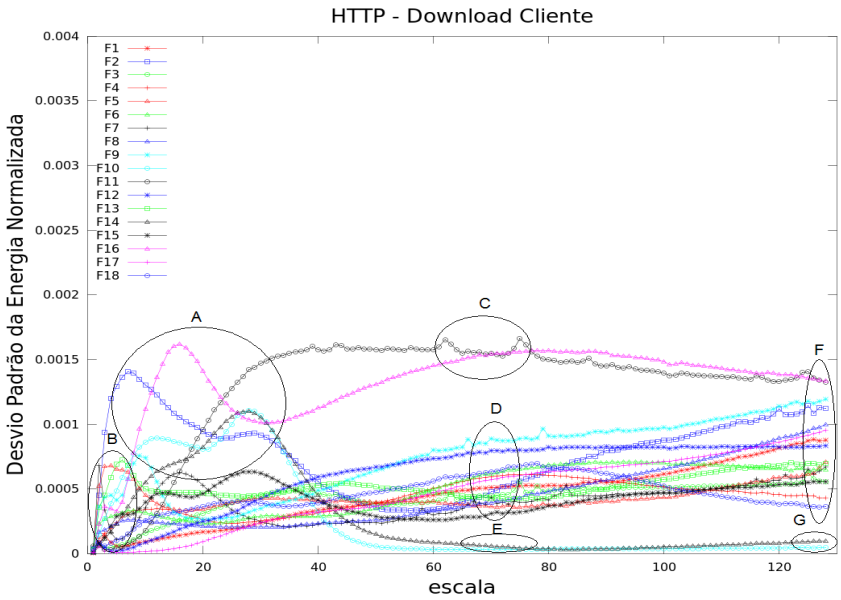

Figura 5.6 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* HTTP (do ponto de vista do cliente).

Portanto, os fluxos contidos nesta região apresentam componentes de alta frequência em quantidade moderada, o que normalmente está associado à chegada de pacotes num volume razoável. Por outro lado, a região G apresenta componentes de alta frequência com variação de energia extremamente reduzida, o que permite deduzir que os dois fluxos (fluxos 10 e 14) são gerados por aplicações responsáveis por tráfego reduzido de pacotes. Estas aplicações poderão ser possivelmente sites de partilha de fotos (não são precisos muitos pacotes para descarregar uma foto) ou clientes de email (nos casos em que os emails têm um tamanho pequeno).

5.1.2 Servidor (*Upstream*)

Esta secção interliga-se com a secção anterior (5.1.1) pois o tráfego capturado que tinha como destino o terminal do utilizador teve a sua origem no servidor, nomeadamente no porto de serviço do protocolo HTTP. O tráfego proveniente dos servidores depende sempre dos pedidos dos clientes, ou seja, quanto maior for a quantidade de pedidos enviados por clientes para o servidor irão ocorrer menos picos de tráfego, pois este será mais constante. Uma menor quantidade de clientes gera tráfego com falhas, ou seja, não é contínuo em todo o intervalo temporal.

Analisando a Figura 5.7, verifica-se a existência de vários picos de tráfego de curta duração e grande amplitude. Os picos com maior amplitude estão relacionados com componentes de baixa frequência mais intensas criadas por cliques dos clientes. Consequentemente, observa-se que o tráfego que surge na sequência dos picos anteriormente referidos apresenta componentes de média e alta frequência, como resposta aos pedidos de vários clientes.

No que diz respeito à Figura 5.8, os picos de tráfego possuem grande amplitude mas com duração um pouco maior comparativamente ao caso da figura anterior. Estão associados a grandes componentes de média e alta frequência, o que alude a transferência de dados. Contudo, tendo em conta que o tráfego em estudo não é contínuo ao longo do tempo, é possível assumir que poucos clientes efetuam pedidos.

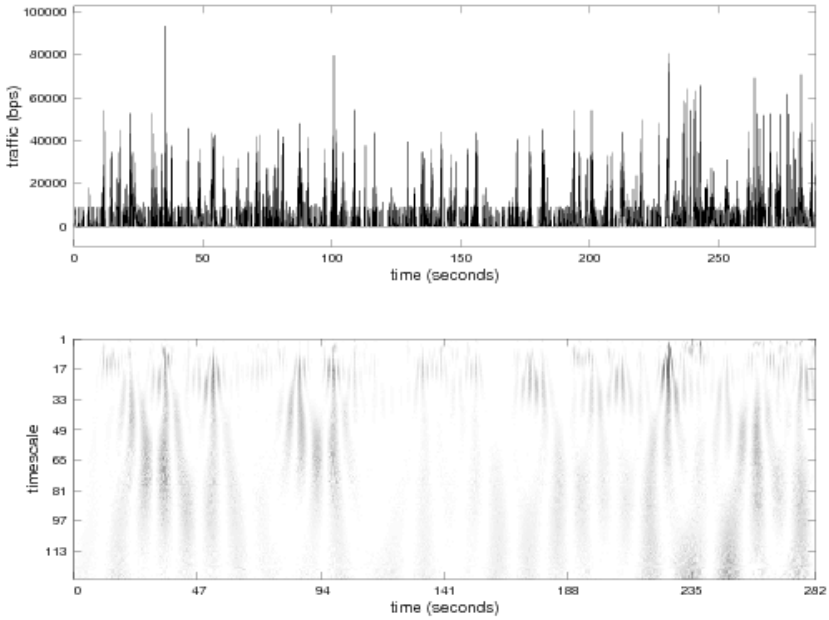

Figura 5.7 – Tráfego *upstream* HTTP por parte do servidor na direção B (bytes por Segundo).

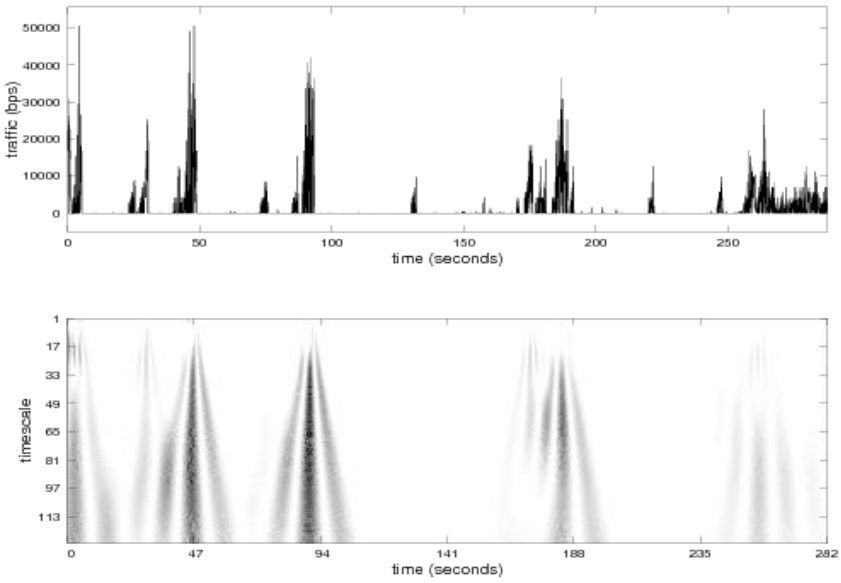

Figura 5.8 - Tráfego *upstream* HTTP por parte do servidor na direção B (bytes por Segundo).

Na Figura 5.9, é possível observar a ocorrência de um pico de tráfego com grande amplitude e com duração de alguns segundos logo no início do intervalo temporal em análise. Este pico de tráfego está associado a componentes de média e alta frequência, o que indicia o início de transferência de dados para o utilizador. De realçar que de seguida surge tráfego contínuo de pacotes de grande tamanho, o que corrobora a ideia de haver transferência de um grande volume de dados neste caso. O tráfego surge com um perfil que não sofre muitas alterações ao longo do tempo, pois é criado pela soma de diferentes fluxos, com picos de tráfego cada, resultando assim no perfil apresentado.

Finalmente, na Figura 5.10 é possível observar a ocorrência de grande volume contínuo de dados, pois existem muitos picos de tráfego com grande amplitude (apesar de possuírem curta duração) que surgem de forma não periódica e são gerados pelos

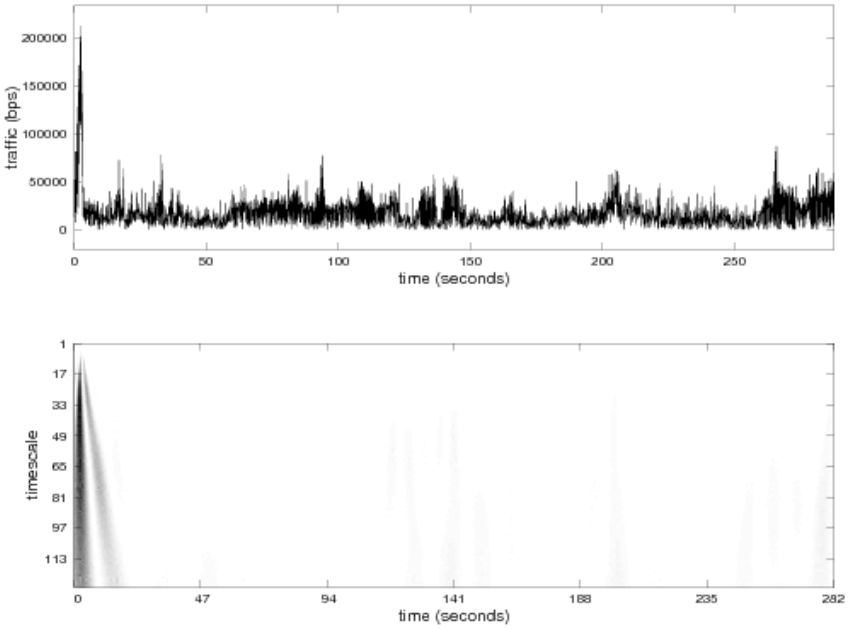

Figura 5.9 - Tráfego *upstream* HTTP por parte do servidor na direção B (bytes por Segundo).

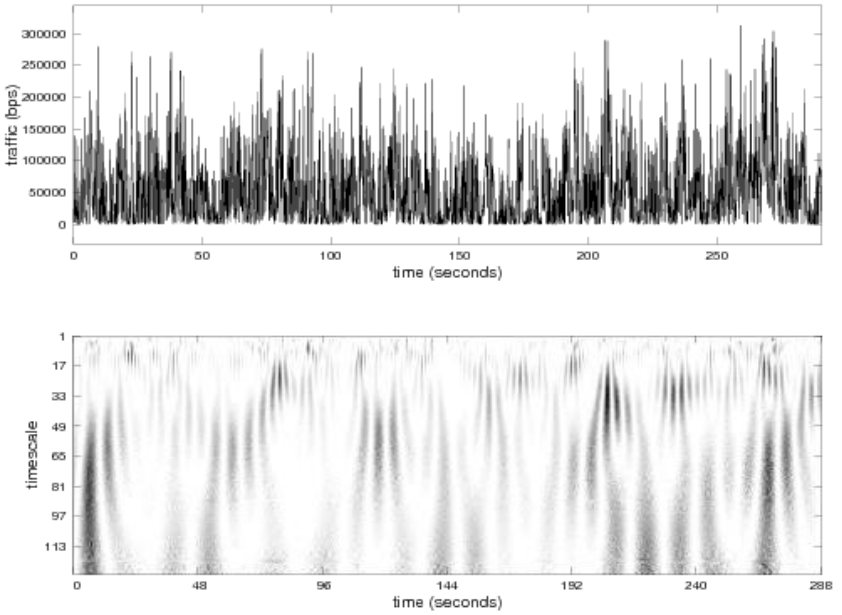

Figura 5.10 - Tráfego *upstream* HTTP por parte do servidor na direção A (bytes por Segundo).

pedidos de vários clientes, que resultam neste perfil de tráfego com múltiplos picos de tráfego. No escalograma surgem componentes de média e alta frequência com bastante relevo e muita regularidade, o que alude à transferência de grande volume de dados. É importante também referir a existência de alguns componentes de baixa frequência, gerados pelas requisições do utilizador. A análise da Figura 5.11é feita de forma similar à análise efetuada à Figura 5.6, mesmo considerando que neste caso o tráfego é visualizado a partir do porto do servidor e é direcionado para os diferentes clientes. A região A situa-se no segmento de baixas frequências e engloba eventos com variação de energia diminuta, logo este tráfego surge como resposta a solicitações pouco frequentes (poucos cliques de utilizador). O segmento de médias frequências tem duas regiões demarcadas. A região B engloba eventos com componentes de média frequência

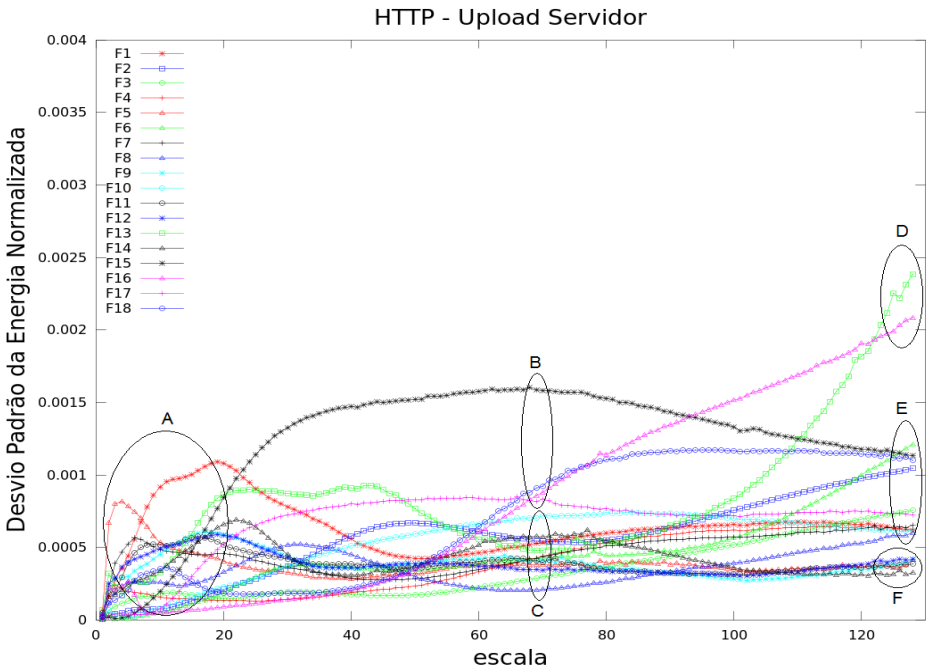

Figura 5.11 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* HTTP (do ponto de vista do servidor).

com moderada variação de energia, o que alude à criação de várias sessões TCP e HTTP em cada fluxo (fluxos 15,16,17 e 18). Já a região C, apesar de estar presente no mesmo segmento de frequências da região B envolve eventos com variação de energia bastante menor. Os eventos com este tipo de características nesta região normalmente estão associados a aplicações de redes sociais, pois as atualizações de estado e a visualização dos *feeds* de notícias não implicam a abertura de muitas sessões. A região D contém dois fluxos (fluxos 13 e 16) e abrange eventos com grande percentagem de componentes de alta frequência, normalmente ligados a aplicações de reprodução e visualização de vídeos. A região E engloba eventos de alta frequência com variação de energia menor comparativamente à região D. Adicionalmente, a região F também se encontra no segmento de altas frequências, mas envolve apenas eventos com variação de energia muito reduzida, como são exemplo a visualização de fotos online (o descarregamento de uma foto requer tráfego diminuto de pacotes) ou a consulta de emails de pequeno tamanho.

5.1.3 Cliente (*Upstream*)

Na situação do envio de tráfego HTTP *upstream* por parte do cliente, encontraram-se três casos distintos. Analisando a Figura 5.12, verifica-se a ocorrência de múltiplos picos de tráfego com elevada amplitude, principalmente no primeiro minuto da janela temporal. Estes picos de tráfego surgem associados a bastantes componentes de média e alta frequência, o que indica a criação de várias sessões TCP e a transferência de uma quantidade razoável de pacotes. O tráfego com estas características relaciona-se com o que se passa no Facebook (upload de uma fotografia ou de um curto vídeo para o mural do cliente, seguido de mensagens enviadas a outras pessoas e criação de *posts* em murais) ou então o que acontece em motores de busca como o Google (escrita de palavras na caixa de busca, seguida dos resultados de busca).

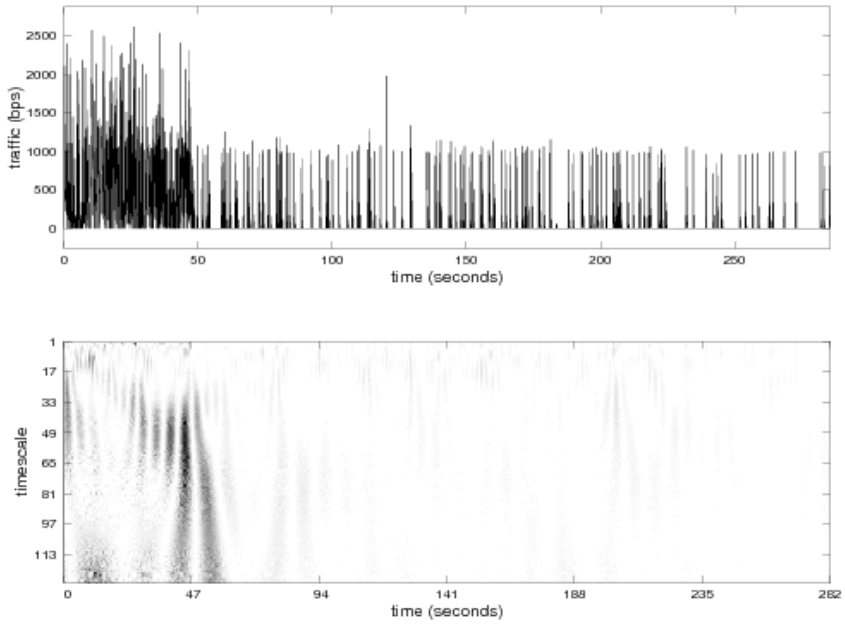

Figura 5.12 – Tráfego *upstream* HTTP por parte do cliente na direção A (bytes por segundo).

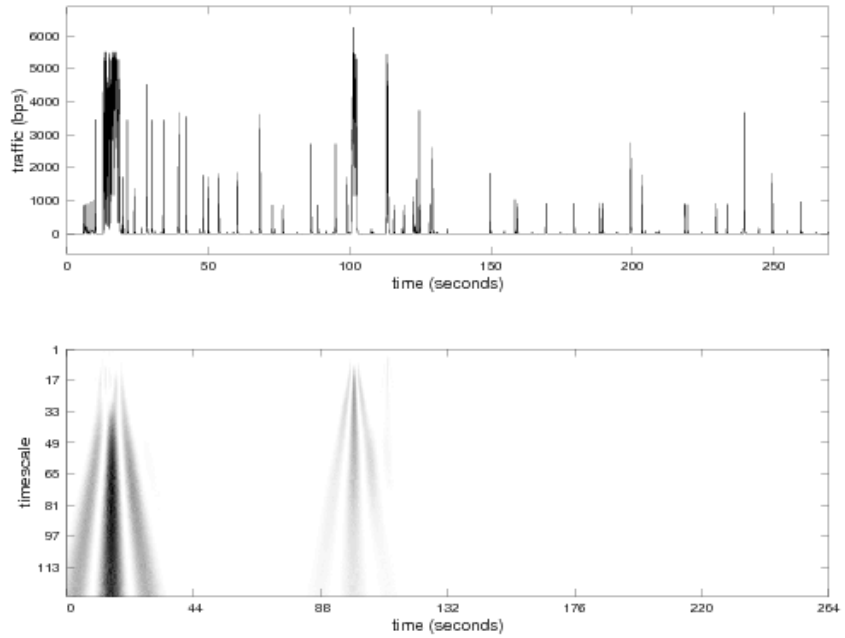

Figura 5.13 - Tráfego *upstream* HTTP por parte do cliente na direção B (bytes por segundo).

A Figura 5.13 e a Figura 5.14constituem dois exemplos típicos de *browsing.* Na Figura 5.13 observa-se que existem menos picos de tráfego comparativamente à figura anterior, havendo dois aglomerados em que estes picos de tráfego não são de curta duração, que correspondem a múltiplos *HTTP Request.* Após estes aglomerados de tráfego iniciais, os *HTTP Request* surgem mais espaçados no tempo. A duração destes picos de tráfego reflete-se no escalograma, pois as componentes de média e alta frequência têm maior amplitude, o que indicia um volume bastante grande de pacotes e portanto o pedido HTTP refere-se a uma página complexa, com mais objetos. As componentes de baixa frequência têm uma presença praticamente residual, devido à pouca intensidade do tráfego. Quanto à Figura 5.14, existe apenas um pico de tráfego em todo o intervalo temporal considerado, de curta duração e com grande amplitude.

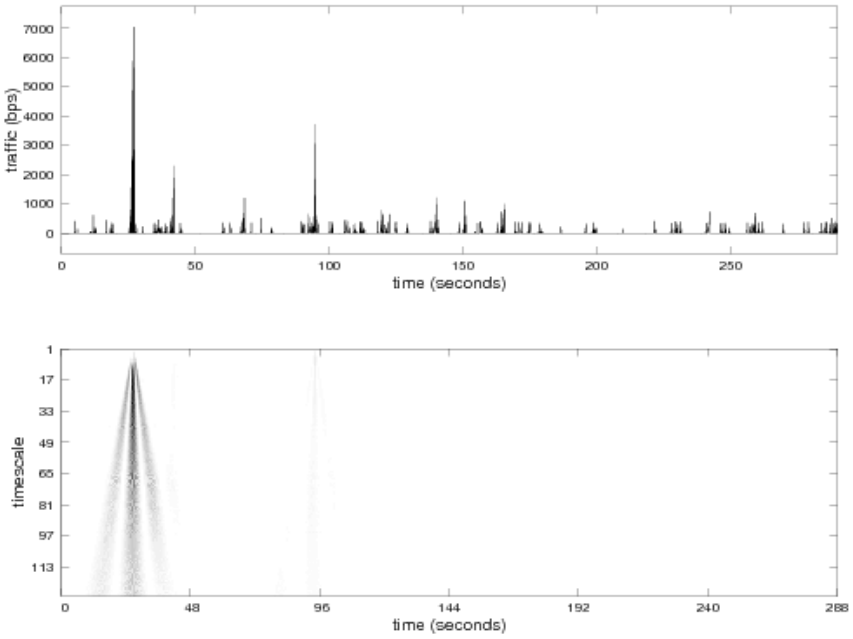

Figura 5.14 - Tráfego *upstream* HTTP por parte do cliente na direção A (bytes por segundo).

No escalograma, a este pico de tráfego surgem associados sobretudo componentes de baixa e média frequência. Portanto, neste caso específico inicialmente surgem múltiplos *HTTP Request*, correspondentes ao pico de tráfego, embora em menor quantidade comparativamente à figura anterior. Ao longo do tempo vão surgindo mais *HTTP Request*, com pouca expressividade e bastante espaçados ao longo do tempo.

A Figura 5.15 apresenta o grafismo de fluxos de tráfego com origem em diferentes clientes e destino no porto destinado ao protocolo HTTP. Analisando esta figura, uma região engloba todos os fluxos no segmento de baixas frequências, sendo que a variação de energia destes fluxos é pequena, o que alude a que estes fluxos sejam originados por cliques pouco frequentes por parte dos vários clientes. No segmento de médias frequências encontram-se dois fluxos (fluxos 10 e 12) responsáveis por eventos com alguma variação de energia (região B); a região C abrange os restantes fluxos de tráfego caracterizados por um número baixo de sessões criadas, donde advém a reduzida variação de energia que estes fluxos apresentam. Estas características geralmente estão relacionadas com aplicações de redes sociais e visualização de vídeos ou fotos. No segmento de altas frequências é possível demarcar duas regiões distintas: a região D contém um fluxo (fluxo 12) caracterizado por uma percentagem razoável de eventos de alta frequência, o que indica que este fluxo é responsável por bastante tráfego *upstream* originário no cliente em causa (como é o caso da Figura 5.13); a região E engloba os restantes fluxos e é caracterizado por eventos com componentes de alta frequência mas com reduzida variação de energia (bastante menor comparativamente à região D), o que aponta para tráfego de pacotes *upstream* reduzido, que é próprio de aplicações em que o utilizador não esteja muito ativo a fazer *browsing* (aplicações de visualização de vídeos, consulta de emails específicos ou consulta dos *feeds* de notícias em redes sociais). Portanto, pode assumir-se que os fluxos representados na Figura 5.12 e Figura 5.14 têm características que os inserem nas regiões A, C e E. Já o tráfego apresentado na Figura 5.13 tem características que o inserem nas regiões B e D, tendo em conta o que foi dito na análise do tráfego dessa figura.

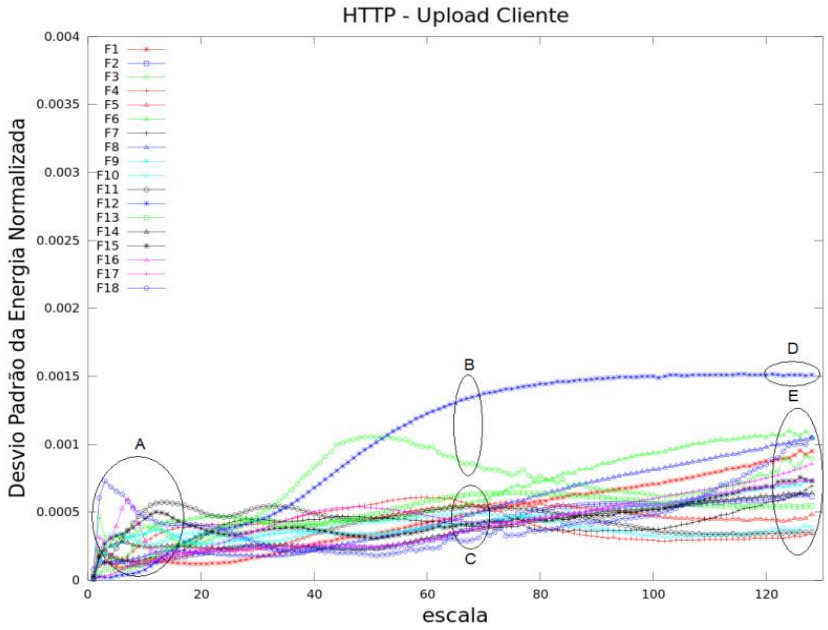

Figura 5.15 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* HTTP (do ponto de vista do cliente).

5.1.4 Servidor (*Downstream*)

Encontraram-se dois casos distintos de tráfego *downstream* HTTP por parte do servidor.

No caso da Figura 5.16, verifica-se a existência de vários picos de tráfego com diferentes características. Os picos de maior amplitude e curta duração estão associados a pequenos componentes de baixa frequência no escalograma. Existem, contudo, picos de tráfego com menor amplitude mas com maior duração, podendo chegar a dezenas de segundos. Ora, estes picos de tráfego estão associados a componentes de média e alta frequência, o que indica que neste caso o cliente está a efetuar upload de conteúdos. Tendo em conta que o tráfego não é contínuo e existem vários momentos em que não há qualquer tipo de tráfego, pode assumir-se que neste caso são poucos os clientes a fazer pedidos ao servidor.

Relativamente à Figura 5.17, verifica-se que existe tráfego *upstream* praticamente constante na direção do servidor, o que aliado à existência de múltiplos picos de tráfego significa que vários clientes efetuam pedidos ao servidor. O facto dos picos de tráfego terem amplitude um pouco elevada indicia que poderão estar a ocorrer transferência de ficheiros dos clientes para o servidor.

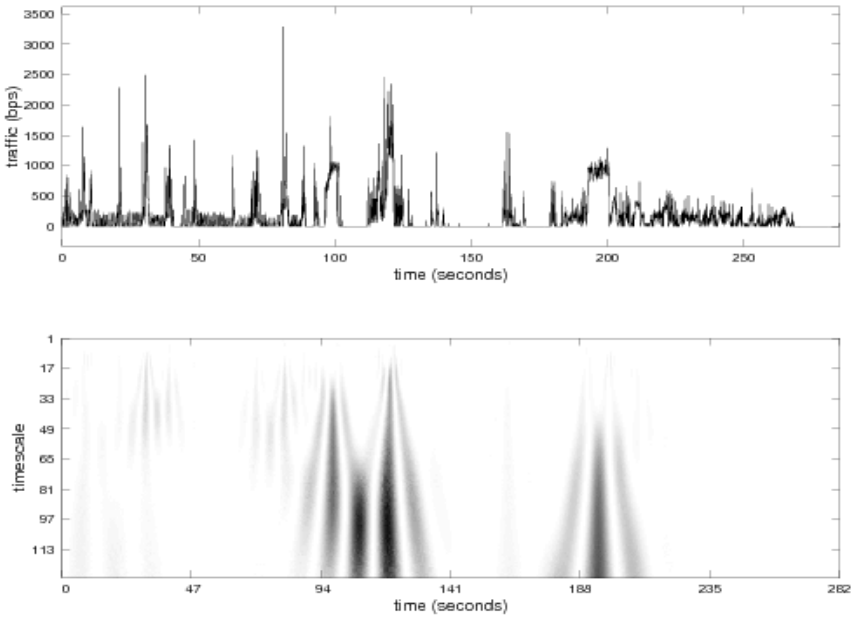

Figura 5.16 - Tráfego *downstream* HTTP por parte do servidor na direção B (bytes por Segundo).

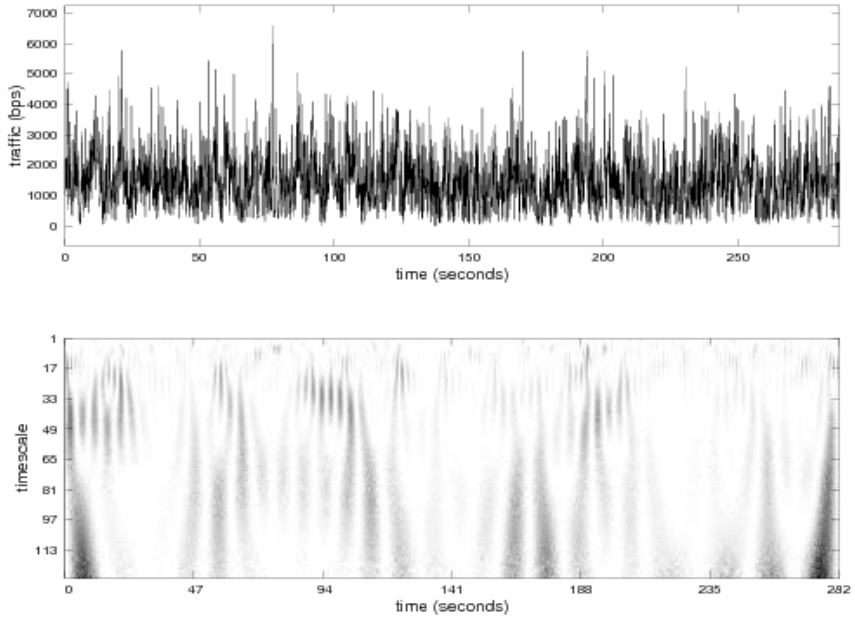

Figura 5.17 - Tráfego *downstream* HTTP por parte do servidor na direção B (bytes por Segundo).

A Figura 5.18 representa o grafismo de fluxos de tráfego *downstream* que têm como destino o porto de serviço do protocolo HTTP, para diferentes endereços IP. A região A engloba eventos de baixa frequência com diminuta variação de energia, envolvendo todos os fluxos considerados neste cenário. No segmento de médias frequências encontram-se duas regiões: a região B congrega tráfego (fluxos 3,4,10,11,13 e 16) com eventos de médias frequências com alguma variação de energia (fluxos de

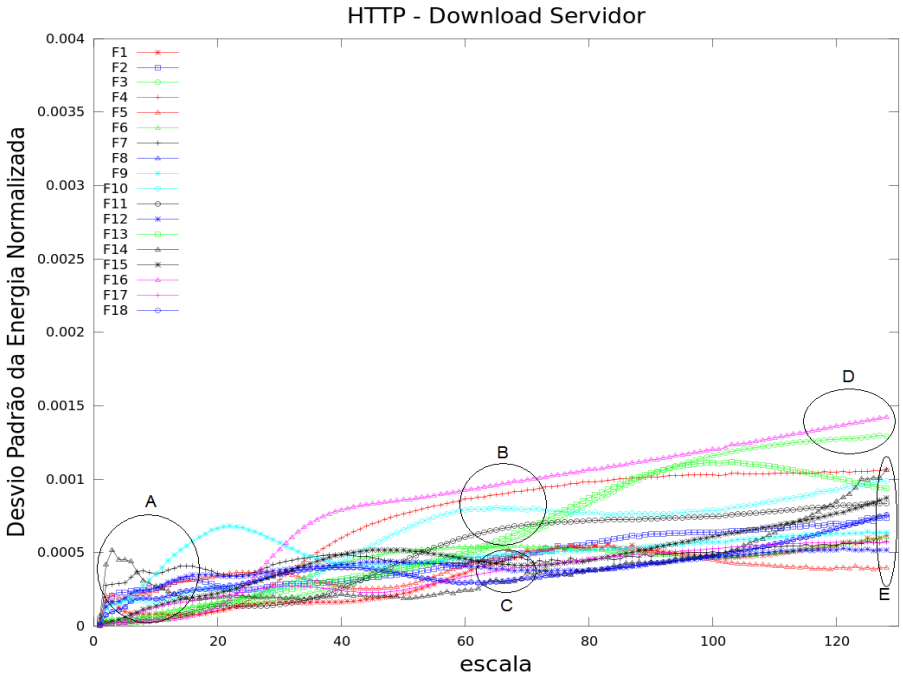

Figura 5.18 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* HTTP (do ponto de vista do servidor).

tráfego gerados pelos clientes relativos a aplicações em que há criação de várias sessões TCP e HTTP); já a região C abrange os fluxos com percentagens reduzidas de componentes de média frequência, devido à criação de poucas sessões e respetiva redução de interações TCP e HTTP, normalmente observáveis em aplicações de redes sociais e consulta de emails. A região D compreende dois fluxos de tráfego (fluxos 3 e 16) com percentagem considerável de componentes de alta frequência, donde resulta um tráfego volumoso de pacotes, comparativamente à região E onde a variação de energia pode ser substancialmente menor. Tendo em conta que os fluxos 3 e 16 estão presentes em ambas as regiões B e D, é possível deduzir que estes fluxos de tráfego são gerados pelos clientes quando estes estão a aceder a sites de consulta de notícias online, por exemplo, pois estes sites requerem a abertura de algumas sessões TCP e envolvem algum volume de pacotes para apresentação da informação pretendida. O tráfego representado na Figura 5.16 apresenta características que o colocam nas regiões A,B e D.

5.1.5 Comparação entre Diferentes Classes de Serviço do Tráfego HTTP

Tendo em conta as características específicas das capturas de tráfego em questão, não é possível fazer uma associação direta sem margem de erro dos fluxos às aplicações que os geraram. Assim, recorreu-se a capturas de tráfego obtidas no ambiente de trabalho referido no Capítulo 4 em que o tráfego analisado foi dividido em cinco classes de serviço: redes sociais (Facebook - www.facebook.com), notícias online (A Bola - www.abola.pt), email (Hotmail – www.hotmail.com), partilha de fotos (Flickr - www.flickr.com) e partilha de vídeos (Youtube – www.youtube.com). O tráfego TCP e HTTP em questão foi capturado num ambiente controlado através do programa *Wireshark*, recorrendo a um computador com ligação de banda larga e sistema operativo Ubuntu. Estas capturas de tráfego foram

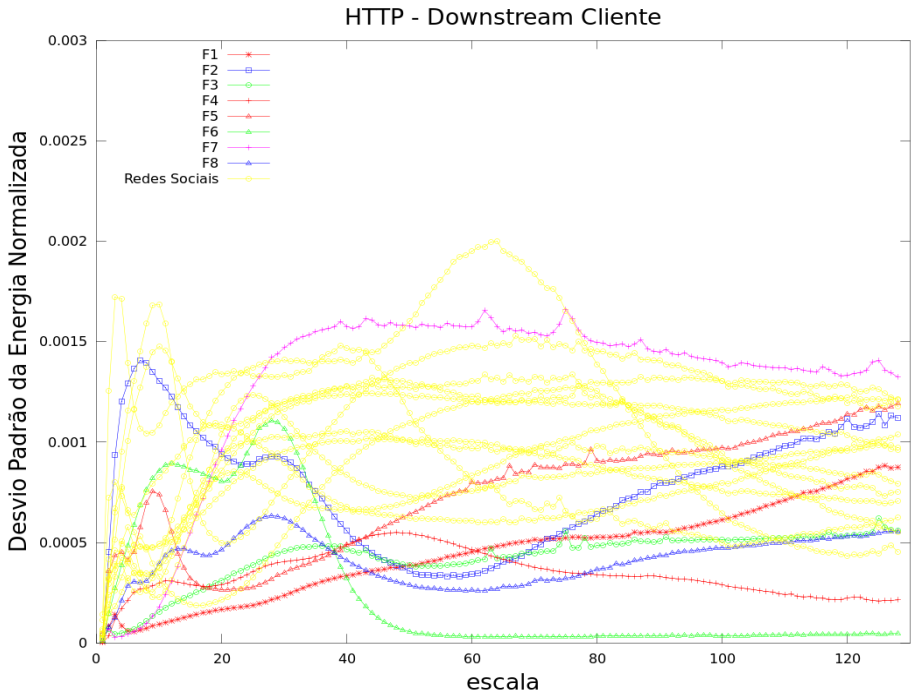

Figura 5.19 – Comparação entre os fluxos de tráfego *downstream* gerados do lado do cliente e fluxos gerados por aplicações de redes sociais.

posteriormente transformadas em tabelas com os dados referentes ao número de bytes e pacotes por intervalo de tempo (0,1 segundos) utilizando o programa *Tshark* (como explicado anteriormente no Capítulo 4). Assim, o tráfego HTTP capturado no âmbito deste trabalho vai ser comparado com cada uma das cinco classes de serviço.

Analisando a Figura 5.19, verifica-se que no segmento de baixas frequências, vários fluxos encontram-se dentro da zona associada aos eventos gerados por aplicações das redes sociais, mas depois apresentam eventos com variação de energia muito baixa no segmento de médias frequências, o que alude a outro tipo de tráfego que crie menos interações TCP e HTTP. Por outro lado, existem fluxos que ao longo de toda a gama de frequências situam-se dentro das zonas associadas às aplicações de redes sociais, portanto poderiam ser elegíveis como tráfego gerado por aplicações de redes sociais.

Relativamente à Figura 5.20, verifica-se que no segmento de baixas frequências os fluxos associados aos eventos gerados pela consulta de notícias online têm variação de energia considerável e dos restantes fluxos apenas o fluxo 2 consegue ter variação de energia suficiente para entrar nessa zona restrita. Como na restante gama de frequências este fluxo situa-se dentro de zonas associadas aos eventos gerados por notícias online, pode assumir-se que este fluxo de tráfego é gerado pela consulta de sites de notícias online ou com características semelhantes.

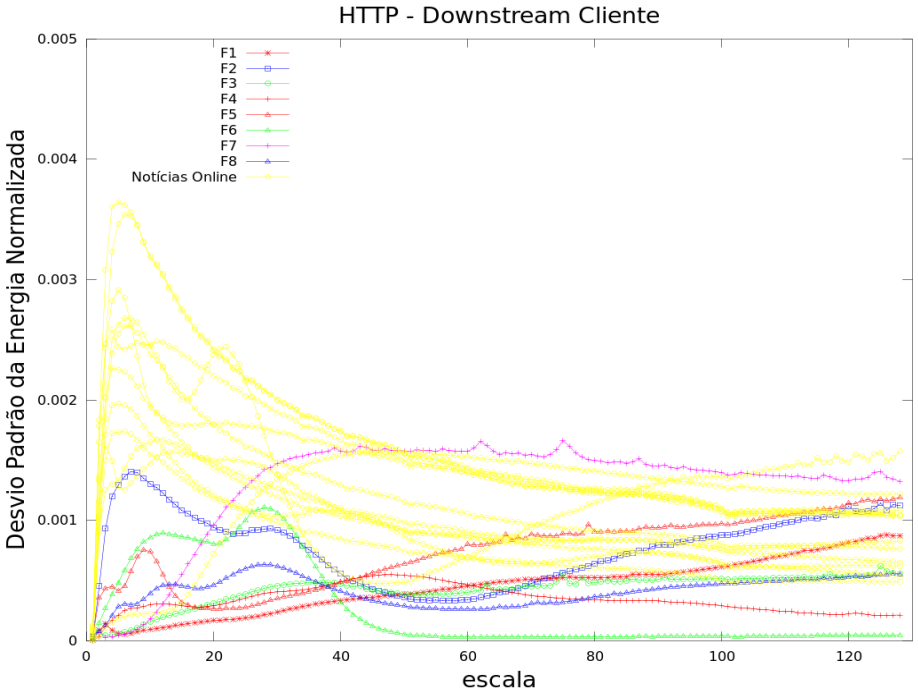

Figura 5.20 - Comparação entre os fluxos de tráfego *downstream* gerados do lado do cliente e fluxos gerados por aplicações de notícias online.

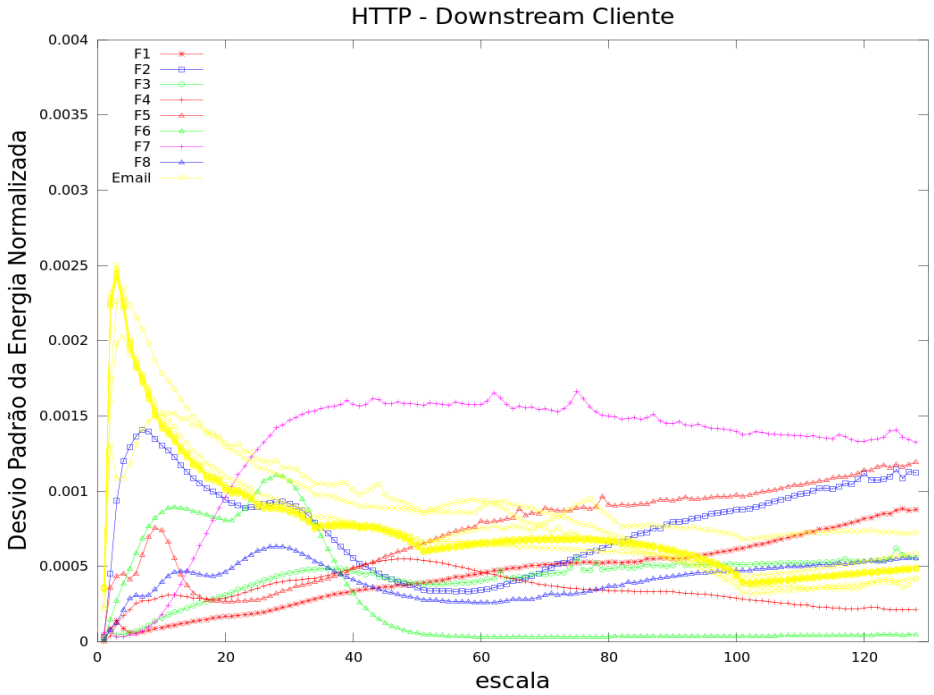

Figura 5.21 - Comparação entre os fluxos de tráfego *downstream* gerados do lado do cliente e fluxos gerados por aplicações de email.

No que concerne à Figura 5.21, verifica-se que no segmento de baixas frequências os eventos gerados por aplicações de email apresentam variação de energia e apenas o fluxo 2 contém componentes de baixa frequência dessa zona. Contudo, no segmento de altas frequências este fluxo mostra uma taxa de transmissão de pacotes

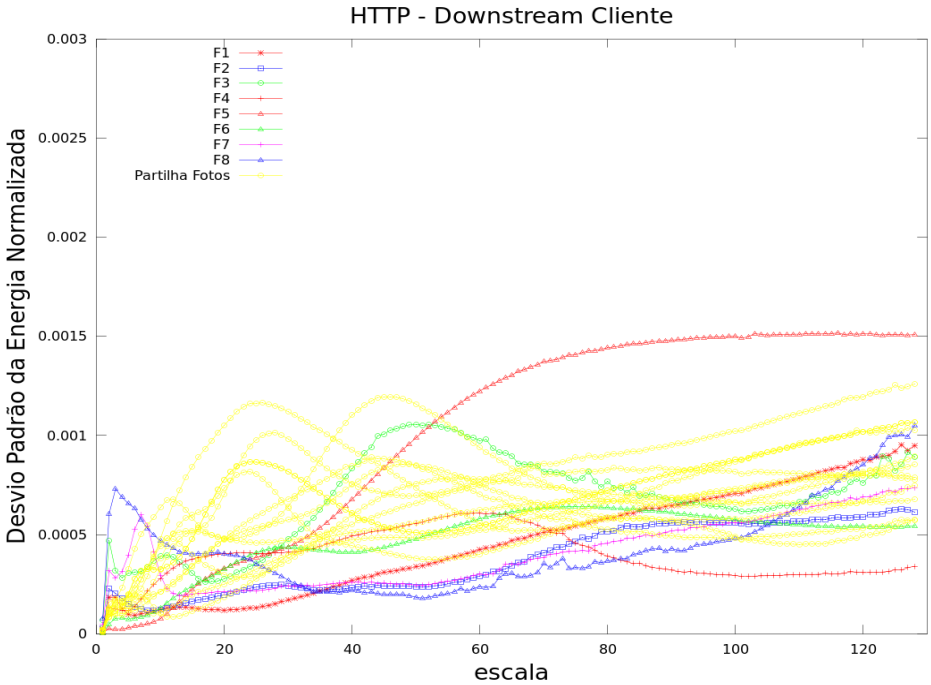

Figura 5.22 - Comparação entre os fluxos de tráfego *downstream* gerados do lado do cliente e fluxos gerados por aplicações de partilha de fotos online.

Superior comparativamente à taxa apresentada pelos fluxos associados à aplicação de email. Tendo em conta que os outros fluxos mostram eventos com uma percentagem de componentes de baixa frequência inferior aos eventos associados à aplicação de email em causa, conclui-se que nenhum dos oito fluxos em análise apresenta características suficientes para se considerar que foi gerado por uma aplicação de email.

A análise da Figura 5.22 permite descortinar que apenas os fluxos 3 e 6 encontram-se em praticamente toda a gama de frequências na zona associada aos eventos gerados por aplicações de partilha de fotos online. Contudo, estas aplicações apresentam características (cliques de utilizador ao efetuar *browsing* geram eventos de baixa frequência; pouca variação de energia no segmento de médias frequências; taxa de transmissão de pacotes no segmento de altas frequências) que podem estar relacionadas com outras aplicações, portanto não é líquido que estes fluxos em análise possam ser associados sem margem de erro a aplicações de partilha de fotos.

Analisando a Figura 5.23, observa-se que apesar de nenhum dos oito fluxos em estudo estar presente em todas os segmentos de frequência na zona associada aos fluxos gerados pela aplicação de vídeo em causa, alguns desses fluxos apresentam um perfil parecido, nomeadamente o fluxo 4 (apesar de ter menor amplitude de variação de energia nas baixas frequências, logo eventos mais raros), o fluxo 2 (apesar de apresentar uma taxa de transmissão de pacotes superior, logo maior transferência de dados) e o fluxo 8 (apesar de ter menor amplitude de variação de energia nas baixas frequências e apresentar uma taxa de transmissão de pacotes superior).

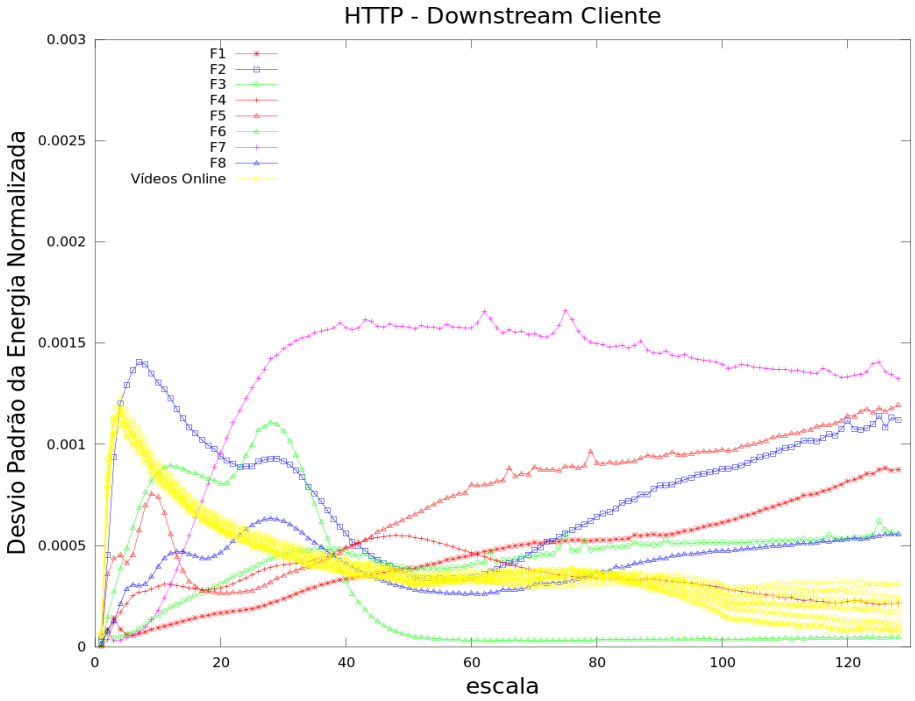

Figura 5.23 - Comparação entre os fluxos de tráfego *downstream* gerados do lado do cliente e fluxos gerados por aplicações de vídeo online.

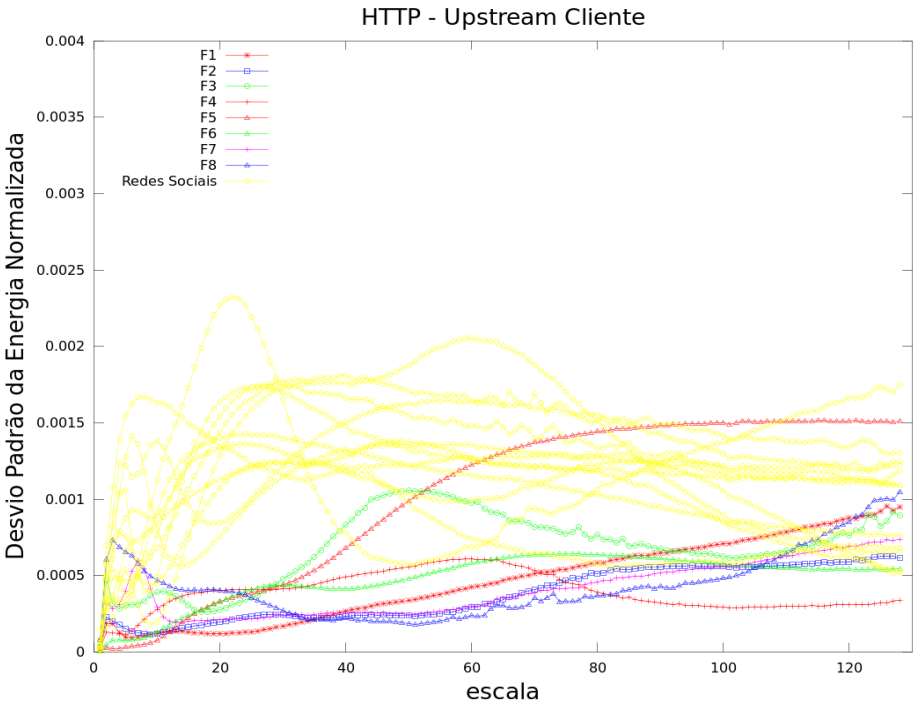

Figura 5.24 - Comparação entre os fluxos de tráfego *upstream* gerados do lado do cliente e fluxos gerados por aplicações de redes sociais.

No que toca aos cenários de tráfego *upstream* HTTP por parte do cliente, estudando a Figura 5.24 verifica-se que apenas o fluxo 3 consegue estar presente em todos os segmentos de frequência nas zonas associadas aos fluxos gerados pela aplicação de redes sociais, enquanto o fluxo 8 tem componentes de baixa e alta

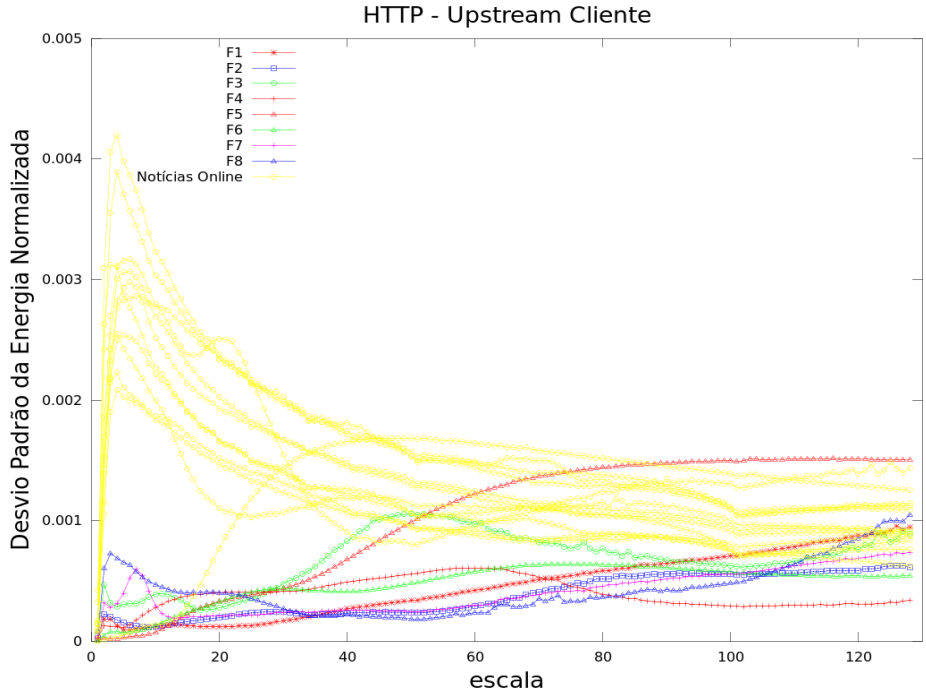

Figura 5.25 - Comparação entre os fluxos de tráfego *upstream* gerados do lado do cliente e fluxos gerados por aplicações de notícias online.

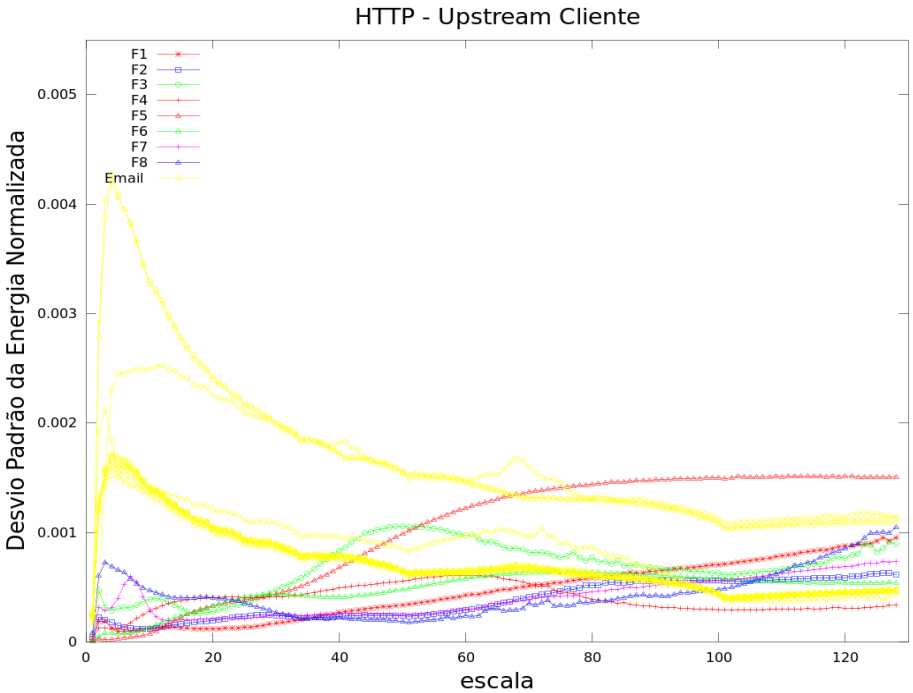

Figura 5.26 - Comparação entre os fluxos de tráfego *upstream* gerados do lado do cliente e fluxos gerados por aplicações de email.

frequência nas zonas demarcadas pelos fluxos da aplicação de redes sociais, mas como mostra pouca variação de energia no segmento de médias frequências, é sinal que cria poucas sessões TCP e portanto não é um fluxo típico de redes sociais.

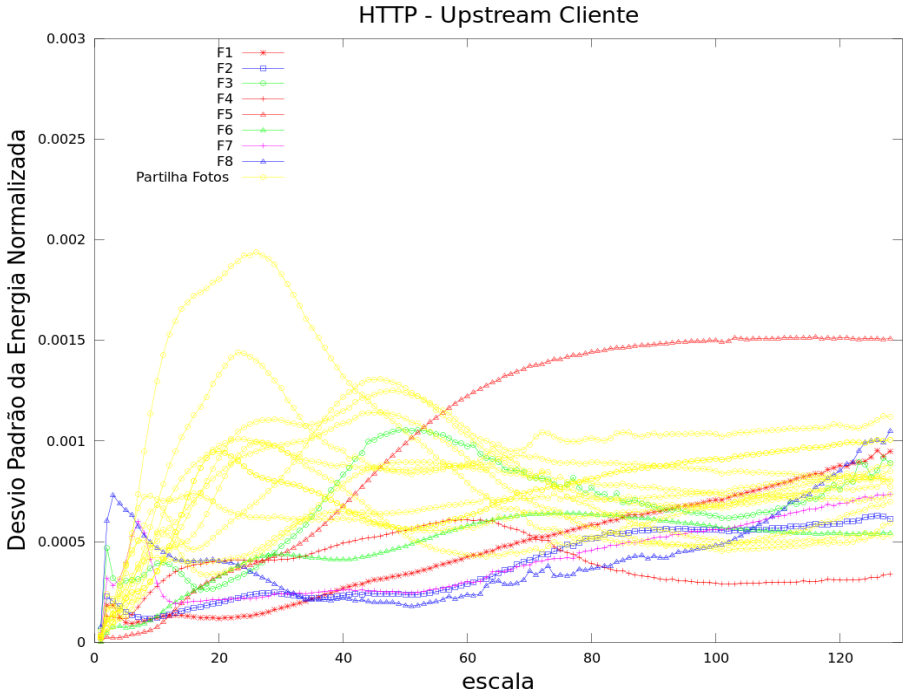

Figura 5.27 - Comparação entre os fluxos de tráfego *upstream* gerados do lado do cliente e fluxos gerados por aplicações de partilha de fotos online.

Analisando a Figura 5.25, observa-se que nenhum dos oito fluxos em estudo apresenta variação de energia como os eventos que constam na zona conotada com a visualização de notícias online. Contudo, os fluxos 3 e 5 mostram percentagens de componentes de média e alta frequência em consonância com as demonstradas pelos fluxos da aplicação de notícias online.

Relativamente à Figura 5.26, observa-se que os fluxos conotados com aplicações de email apresentam eventos de muito baixa frequência com grande variação de energia no tráfego *upstream* HTTP, o que não é o caso de nenhum dos oito fluxos em estudo.

No que concerne à Figura 5.27, verifica-se que vários fluxos encontram-se na zona associada às aplicações de partilha de fotos online ao longo de toda a gama de frequências, mas como referido anteriormente este tipo de aplicação apresenta características em comum com outras aplicações de *browsing* e como tal existe uma certa margem de erro em classificar estes fluxos como sendo gerados por aplicações de partilha de fotos.

Finalmente, a análise da Figura 5.28 permite verificar que, tal como na Figura 5.23, nenhum dos oito fluxos em estudo está presente em todas os segmentos de frequência na zona associada aos fluxos gerados pela aplicação de vídeo em causa. Porém, alguns desses fluxos apresentam um perfil parecido, nomeadamente os fluxos referidos na Figura 5.23 (fluxos 2, 4 e 8).

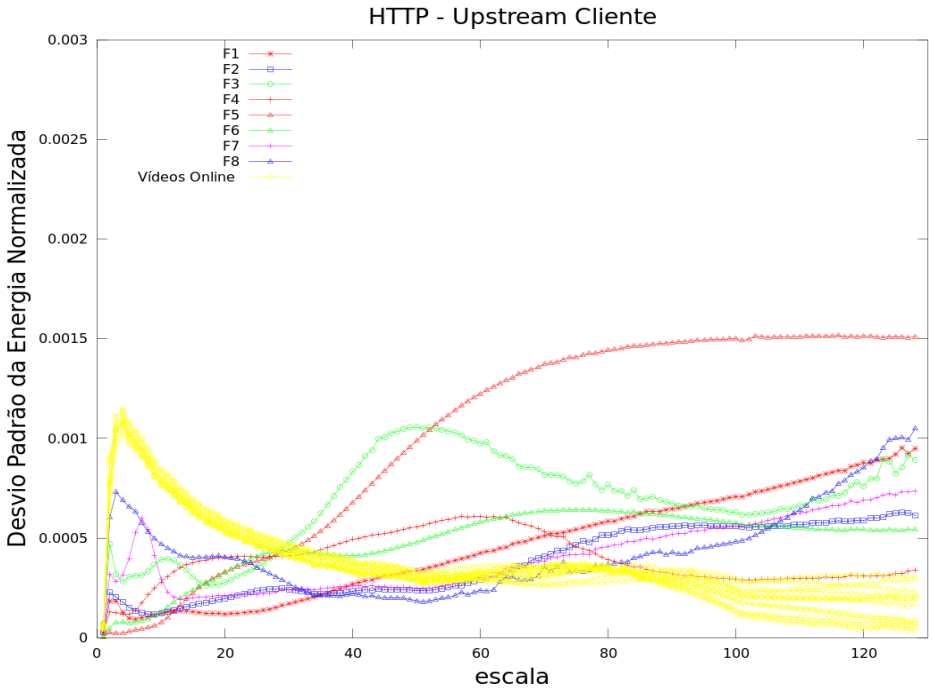

Figura 5.28 - Comparação entre os fluxos de tráfego *upstream* gerados do lado do cliente e fluxos gerados por aplicações de vídeo online.

5.2 SMTP

5.2.1 Cliente (*Downstream*)

Para a situação do tráfego *downstream* SMTP por parte do cliente foram encontrados três casos distintos.

Na Figura 5.29, o tráfego que o cliente recebe surge em três agregados, de duração variada. Não existem picos de tráfego, o que explica a ausência de componentes de baixa frequência no escalograma. Existem apenas componentes relevantes de alta frequência no início do intervalo de tempo em estudo, gerados pelo tráfego de sincronização que verifica se existem novos emails. A ausência de componentes de baixa frequência sugere que este tráfego seja SPAM ou resulte da comunicação entre servidores, pois as características deste tráfego indicam a ocorrência de download de um ficheiro grande ou de vários ficheiros de uma só vez sem que o cliente o tenha solicitado.

No que concerne à Figura 5.30, ocorreram alguns picos de tráfego, de curta duração e pequena amplitude. A elevada presença de componentes de média e alta frequência estão ligadas à criação de várias sessões TCP e transferência de pacotes, ou seja, pode indiciar que este tráfego refere-se à chegada dos pacotes de confirmação (*acknowledge*) do envio de emails ou outros conteúdos por parte do utilizador.

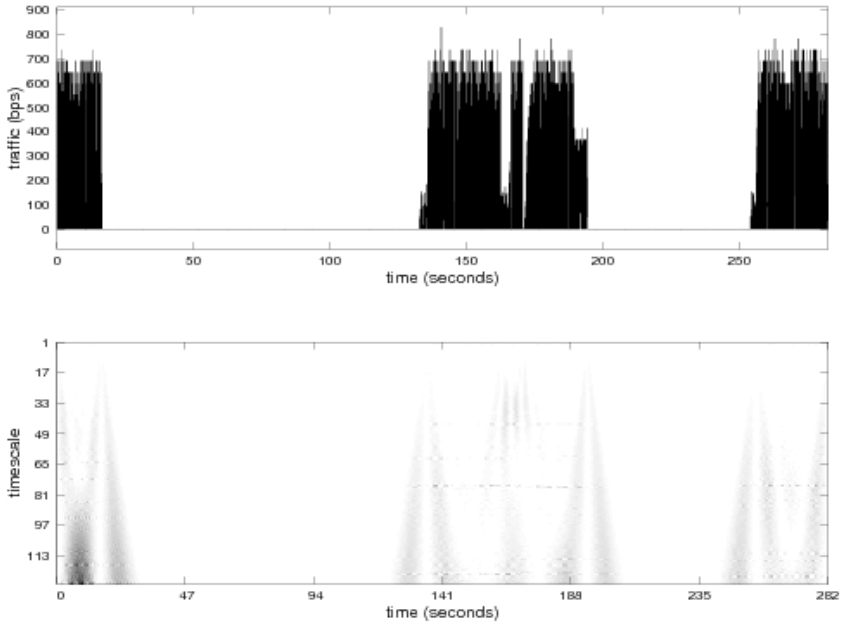

Figura 5.29 - Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

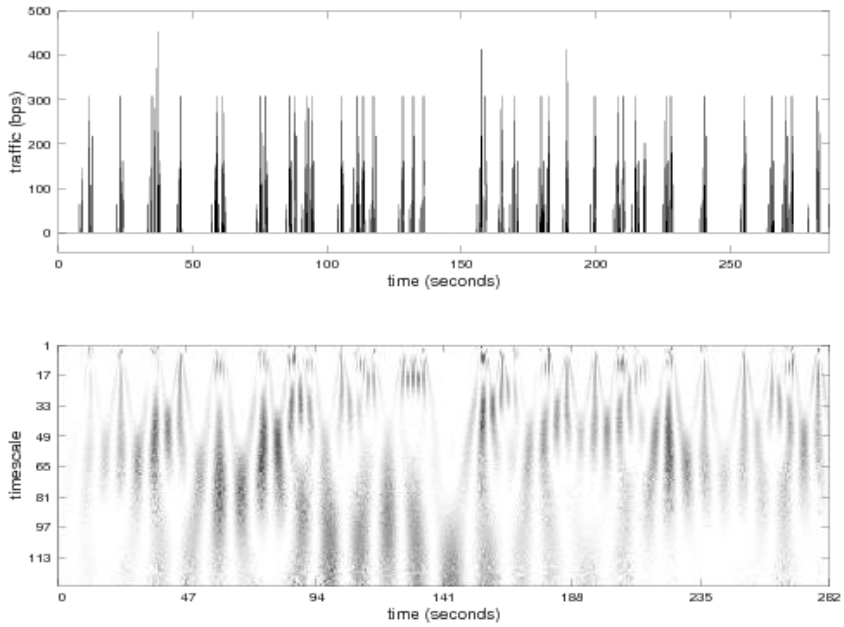

Figura 5.30 - Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

Finalmente, relativamente à Figura 5.31 verifica-se a ocorrência de três picos de tráfego, sendo que um deles tem duração e amplitude superior aos outros dois. Esse pico de tráfego, que surge por volta dos duzentos segundos está associado às componentes de baixa, média e alta frequência presentes no escalograma. A grande amplitude das componentes de média e alta frequência dá a entender a criação de uma sessão por parte do utilizador, o qual recebe os pacotes vindos do servidor nesta altura.

A análise à Figura 5.32 é efetuada de forma similar aos gráficos representativos do desvio padrão da energia normalizada dos fluxos de tráfego analisados em cada

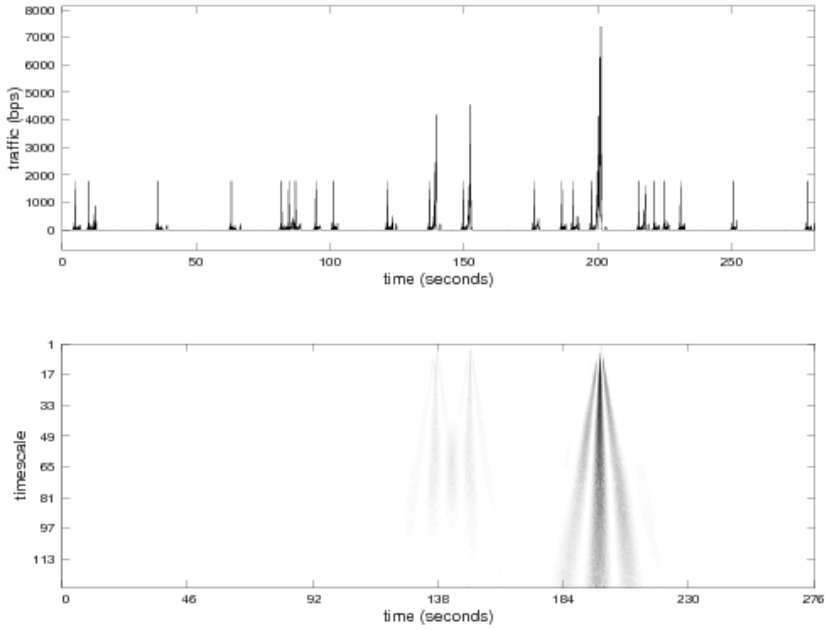

Figura 5.31 - Tráfego *downstream* SMTP por parte do cliente na direção B (bytes por segundo).

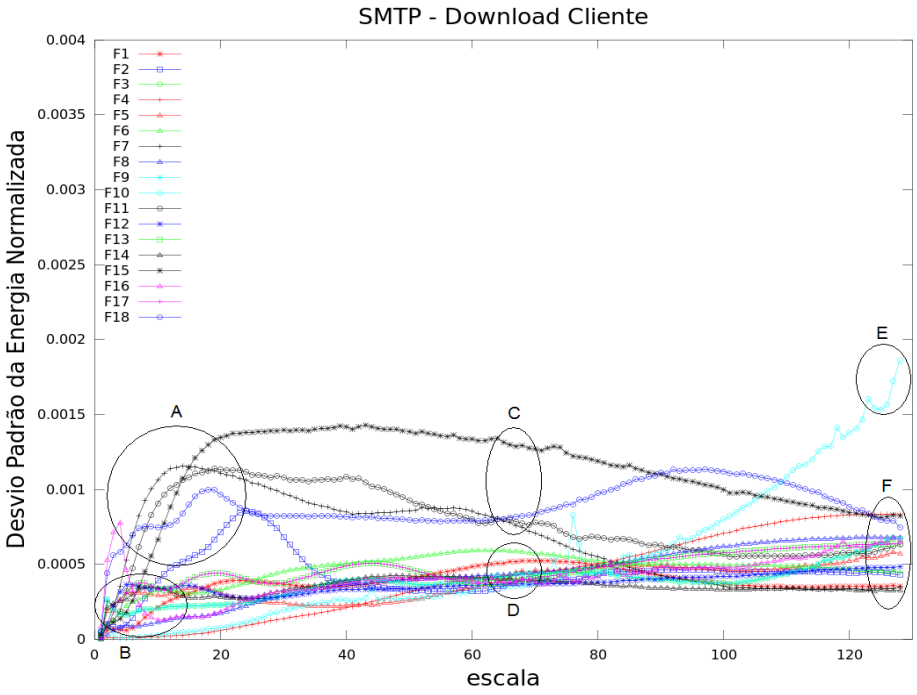

Figura 5.32 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* SMTP (do ponto de vista do cliente).

cenário na secção 5.1. Assim, analisando esta figura é possível verificar que a grande parte dos fluxos de tráfego deste cenário assumem um comportamento semelhante, estando incluídos em três regiões de diferentes segmentos de frequência em simultâneo: regiões B, D e F. A sua presença nestas regiões permite assumir que estes fluxos de tráfego apresentam uma percentagem baixa de componentes de frequência ao longo de toda a escala, associadas a poucos cliques do utilizador, abertura de poucas sessões

TCP e transporte de pacotes em quantidades reduzidas. Estes dados estão de acordo com o expectável para o download de tráfego por parte do cliente neste protocolo. Contudo, existem alguns fluxos que fogem a este padrão. Na região A (fluxos 2,7,11,15 e 18) encontram-se eventos de baixa frequência com variação de energia moderada, o que implica mais cliques do utilizador comparativamente aos fluxos encontrados na região B. A região C (fluxos 7,11,15 e 18) congrega eventos de média frequência com variação de energia moderada, o que indica que nestes fluxos há mais interações HTTP e TCP comparativamente ao que sucede na região D. Nestes casos há mais páginas a serem abertas, resultantes de um maior tráfego de emails. Finalmente, na região E encontra-se apenas um fluxo com variação de energia bastante grande (fluxo 10), o que indicia que neste caso há envio de um ou mais emails, tendo em conta o tráfego volumoso de pacotes detetado.

5.2.2 Servidor (*Upstream*)

A Figura 5.33 e a Figura 5.34 são dois casos em que o tráfego *upstream* SMTP para o servidor é contínuo e não periódico. Ambos têm picos de tráfego com grande amplitude mas de curta duração assim como picos de tráfego de maior duração mas com amplitude menor comparativamente aos primeiros. Isto indica que estamos na presença de alguns clientes a efetuar pedidos ao servidor; tendo em conta a quantidade de picos de tráfego, o número de pedidos não é muito elevado. Os picos com maior amplitude normalmente estão associados a componentes de baixa frequência no escalograma (pedidos do cliente) enquanto os picos de tráfego com maior duração estão associados a componentes de média e alta frequência, o que poderá implicar a existência de transferência de ficheiros (neste caso envio de emails) ou atualização da caixa de correio do utilizador.

Relativamente à Figura 5.35, existe um pico de tráfego de pequena amplitude e com duração de alguns segundos, no primeiro minuto da amostra. Está relacionado com componentes de média e alta frequência no escalograma o que indicia que haja envio de emails. Os restantes picos de tráfego têm pequena amplitude, portanto serão essencialmente pacotes de controlo da ligação ou relacionados com pequenas operações na caixa de correio. Tendo em conta a pouca amplitude do tráfego e volume baixo, nesta situação existem poucos clientes que efetuam poucos pedidos, daí o tráfego ser tão baixo.

Analisando a Figura 5.36, é possível verificar que a maior parte dos fluxos de tráfego considerados apresentam comportamento semelhante ao perfil de comportamento padrão dos fluxos de tráfego representados na Figura 5.32 na situação do tráfego *downstream* SMTP para o cliente. Neste caso em análise, encontram-se fluxos de tráfego em que o servidor procede à abertura de várias sessões TCP a pedido do cliente, como se presencia na região B no segmento de médias frequências. Observa-se também a presença de tráfego na região D (fluxos 10,12,16 e 17) do segmento de altas frequências com uma variação de energia superior comparativamente ao tráfego encontrado na região E, devido a trocas em maior quantidade de pacotes entre servidor e cliente, o que indicia que nestes casos o cliente efetuou mais pedidos ao servidor ao consultar a sua caixa de correio eletrónico.

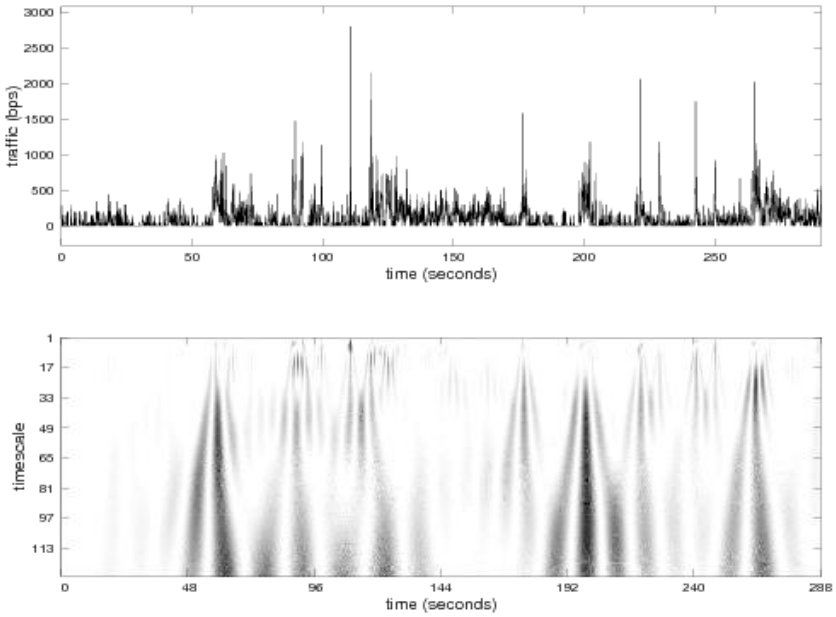

Figura 5.33 - Tráfego *upstream* SMTP por parte do servidor na direção A (bytes por segundo).

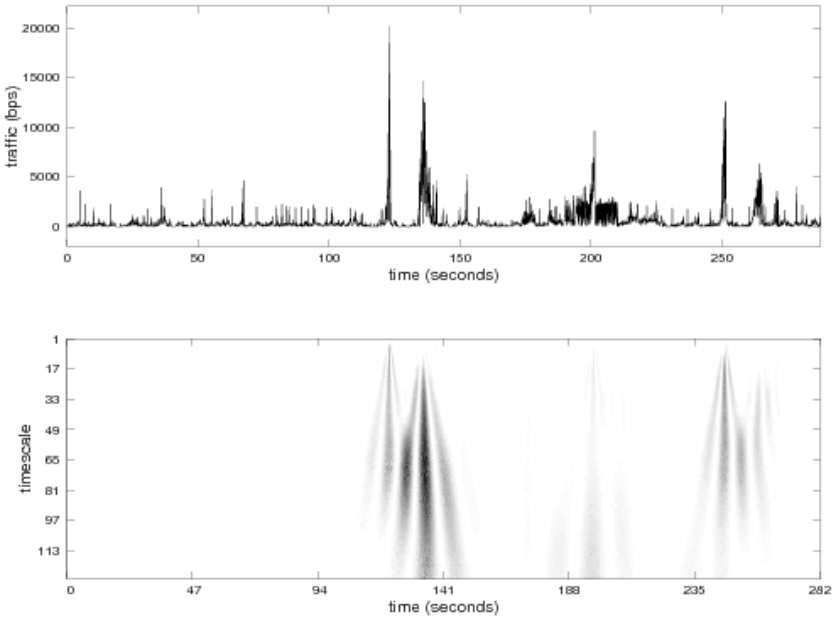

Figura 5.34 - Tráfego *upstream* SMTP por parte do servidor na direção B (bytes por segundo).

As características dos fluxos de tráfego apresentados na Figura 5.33, Figura 5.34, e Figura 5.35 permitem assumir que a variação de energia destes fluxos ao longo dos diferentes segmentos de frequência estará na gama dos fluxos que se encontram nas regiões A, C e E.

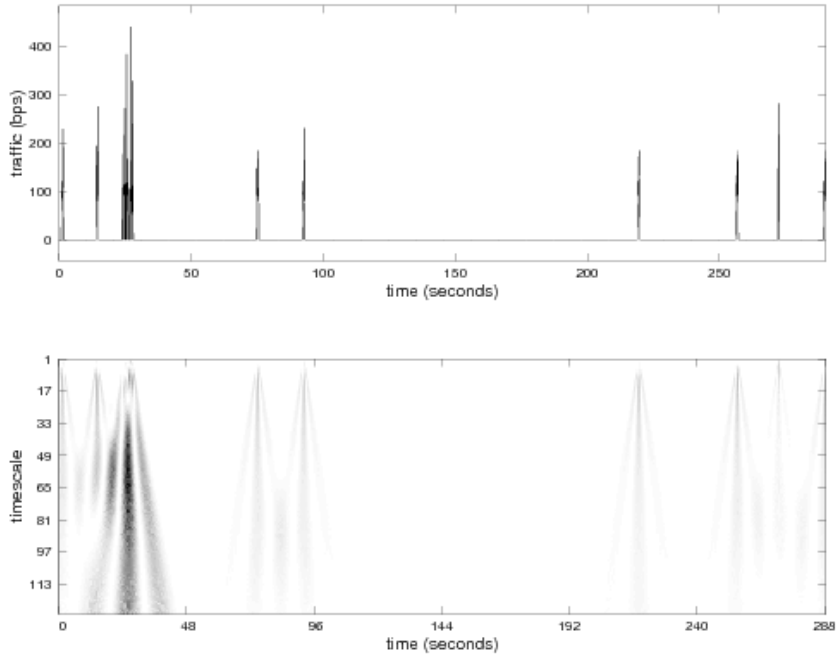

Figura 5.35 - Tráfego *upstream* SMTP por parte do servidor na direção A (bytes por segundo).

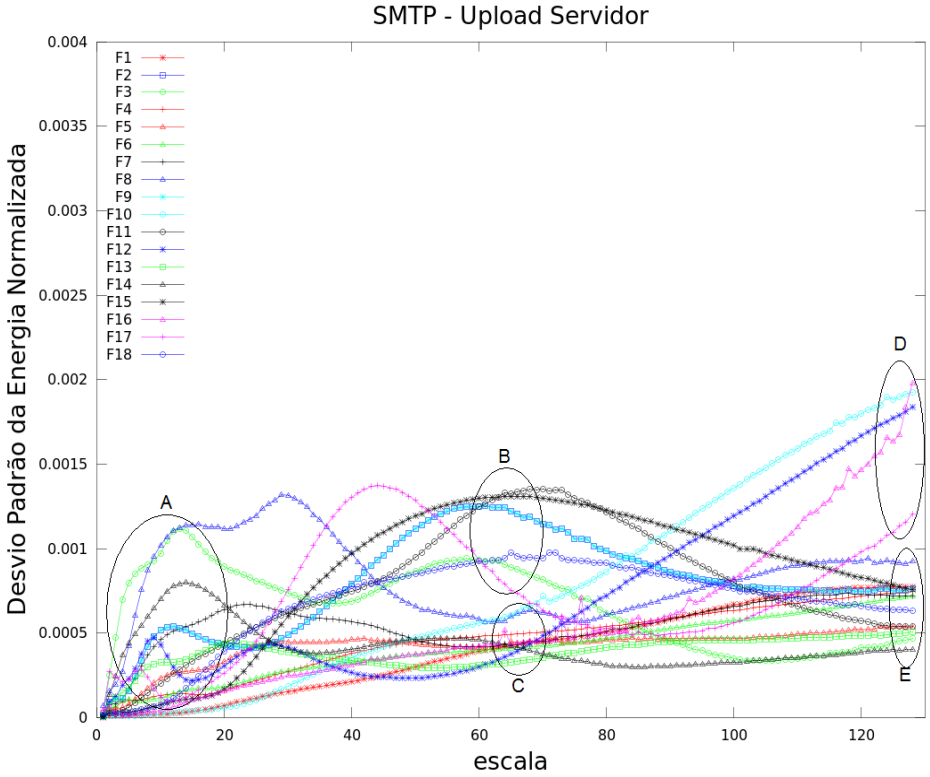

Figura 5.36 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* SMTP (do ponto de vista do servidor).

5.2.3 Cliente (*Upstream*)

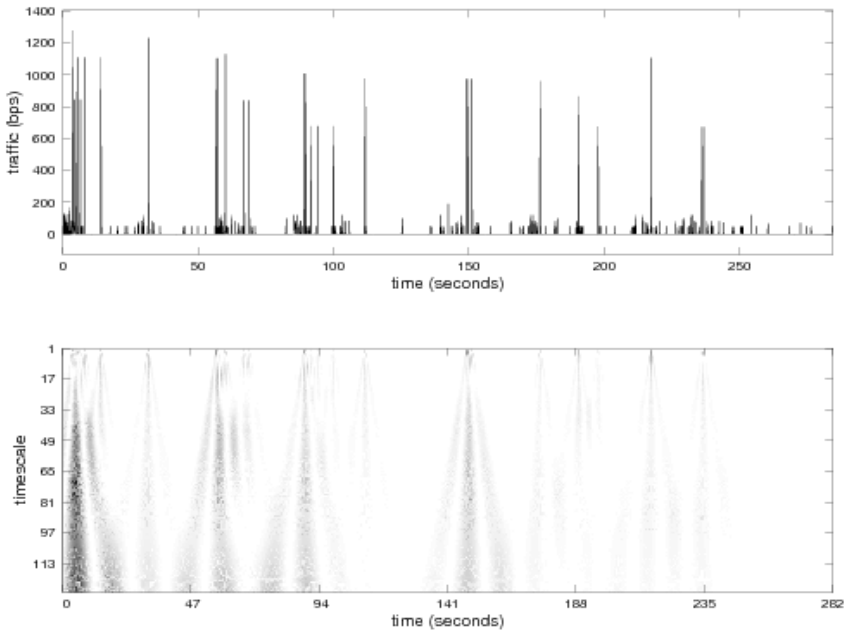

Figura 5.37 - Tráfego *upstream* SMTP por parte do cliente na direção B (bytes por segundo).

Para a situação do tráfego *downstream* SMTP por parte do cliente encontraram-se dois casos diferentes. No caso da Figura 5.37, encontram-se vários picos de tráfego de média amplitude e curta duração. Contudo, no início do intervalo de amostragem surgem picos de tráfego de maior duração, o que implica componentes de média e alta frequência mais salientes e de maior amplitude comparativamente aos associados aos outros picos de tráfego. Pode então dizer-se que no início desta amostra o utilizador criou uma sessão, ligando-se ao servidor. Os outros componentes de alta frequência estarão ligados preferencialmente à interação do utilizador com a caixa de correio.

Relativamente à Figura 5.38, verifica-se a existência de poucos picos de tráfego, de curta duração e baixa amplitude, que correspondem à sincronização automática que é efetuada aquando do estabelecimento da ligação entre o utilizador e o servidor. Sensivelmente a partir do terceiro minuto da janela temporal começam a circular pacotes de controlo de pequeno tamanho, registando-se porém a ocorrência de picos de tráfego ocasionais. Os picos de tráfego desta amostra, sendo de curta duração, originam sobretudo pequenos componentes de média e alta frequência, resultantes do tráfego de sincronismo que verifica o estado das caixas de correio em termos de movimentação de emails.

Analisando a Figura 5.39, existem duas regiões no segmento de baixas frequências: a região A engloba eventos de frequência muito baixa, gerados pelo download inicial por parte do cliente da interface da aplicação de email e posteriores sincronizações automáticas e atualizações da caixa de correio eletrónico do cliente. Assim, a região B corresponde às situações em que além destes eventos, o cliente efetua outras operação na sua interface de email, nomeadamente envio de emails para outros contactos. A região C compreende todos os fluxos de tráfego no segmento de médias frequências, com variação de energia pequena ou moderada, patenteando a criação de

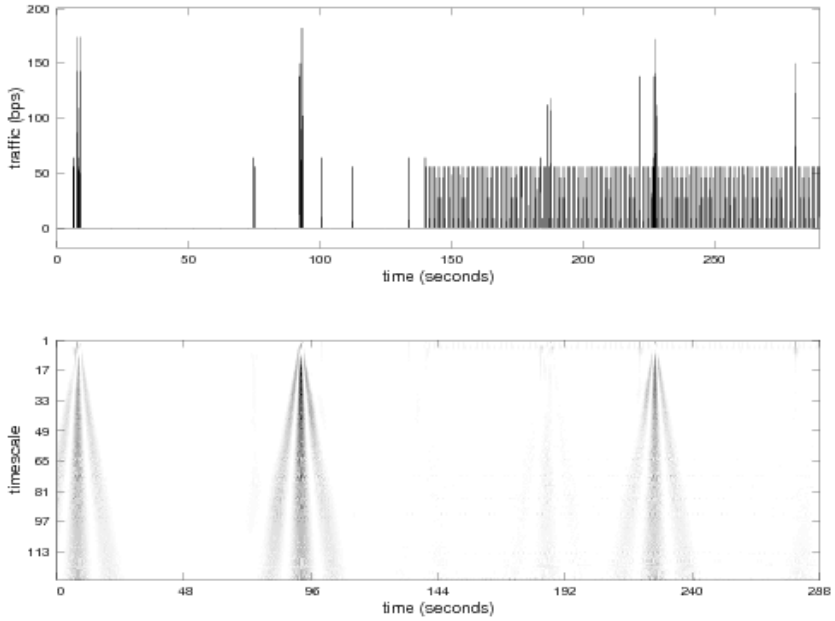

Figura 5.38 - Tráfego *upstream* SMTP por parte do cliente na direção A (bytes por segundo).

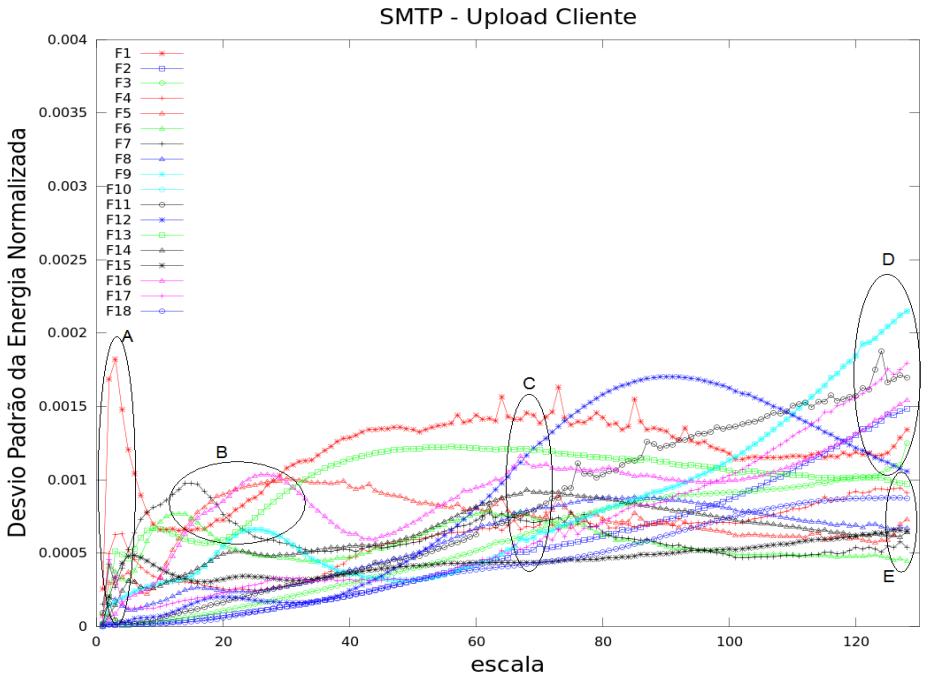

Figura 5.39 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* SMTP (do ponto de vista do cliente).

algumas sessões TCP durante o intervalo de tempo de amostragem. Finalmente, no segmento de altas frequências encontram-se duas regiões: região D, com grande percentagem de componentes de alta frequência e região E, com percentagem reduzida de componentes de alta frequência. A região D estará assim associada a eventos gerados pelo envio de pacotes em grandes quantidades do cliente para o servidor (nomeadamente envio de emails), enquanto a região E estará relacionada com situações em que a atividade do cliente seja mais reduzida e haja apenas troca de pacotes de controlo e sincronização entre o interface da aplicação de email e o servidor.

5.2.4 Servidor (*Downstream*)

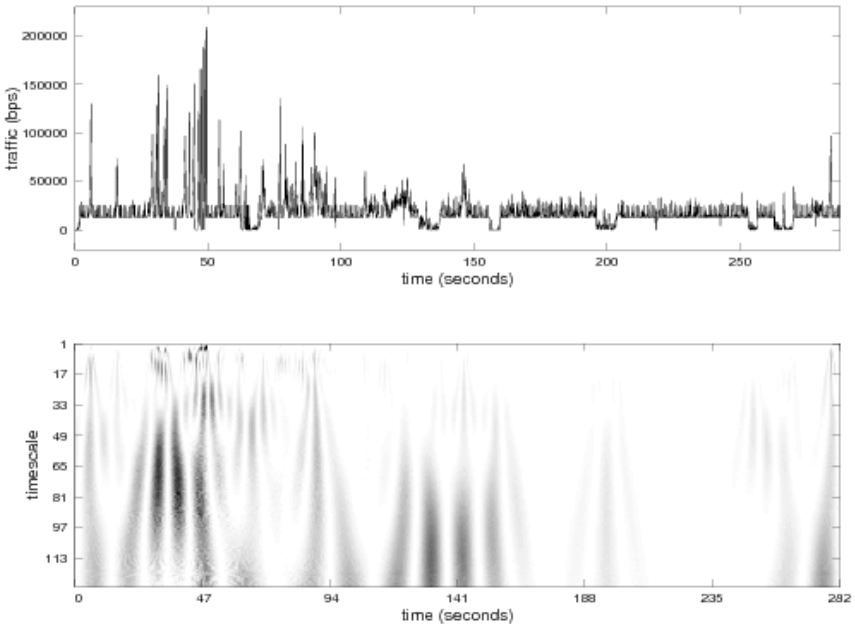

Figura 5.40-Tráfego *downstream* SMTP por parte do servidor na direção B (bytes por segundo).

A análise das seguintes três figuras (Figura 5.40, Figura 5.41 e Figura 5.42) permite encontrar um ponto em comum: durante o intervalo temporal considerado em cada um dos casos é possível encontrar um pico de tráfego de moderada duração, associado a componentes de alta frequência no escalograma.

Na Figura 5.40, o pico de tráfego inicial cria componentes de média frequência, iniciando assim a sessão TCP de modo a ser possível a troca de ficheiros que posteriormente acontece. A transferência de conteúdos não é limitada ao pico de tráfego referido no parágrafo anterior, pois este é sucedido por vários picos de tráfego de amplitude mais baixa mas com componentes de alta frequência relevantes. A continuidade do tráfego e o surgimento de alguns picos de tráfego de curta duração ao longo do tempo sugerem que alguns clientes efetuaram pedidos ao servidor e estão ativos durante grande parte do intervalo temporal considerado. Na Figura 5.41 as componentes de alta frequência apenas surgem aquando do primeiro pico de tráfego, pois os seguintes picos de menor amplitude não geram componentes de frequência observáveis no escalograma. Tendo em conta que em grande parte do tempo não existe tráfego, pode assumir-se que poucos clientes estão ativos neste cenário (possivelmente apenas um) e enviam poucos pedidos ao servidor; neste cenário o pico de tráfego mais acentuado corresponde ao envio de um email de grande tamanho (ou vários emails pequenos). A Figura 5.42 apresenta tráfego que não encaixa no perfil esperado para o tráfego *downstream* SMTP com destino ao servidor, por isso tendo em conta os picos de tráfego consecutivos com duração considerável este tráfego poderá ser SPAM (pois estes picos de tráfego estão associados a componentes de baixa frequência) ou então comunicação entre servidores.

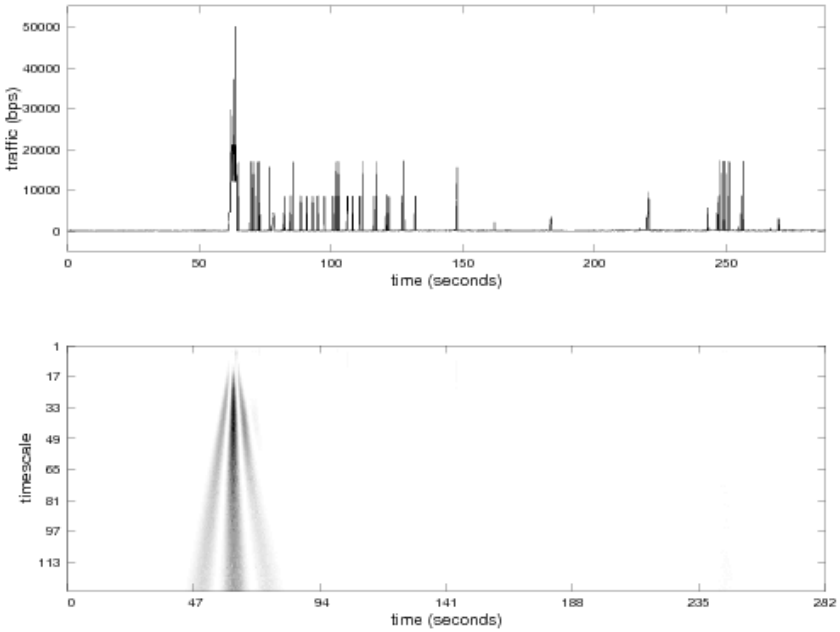

Figura 5.41 - Tráfego *downstream* SMTP por parte do servidor na direção B (bytes por segundo).

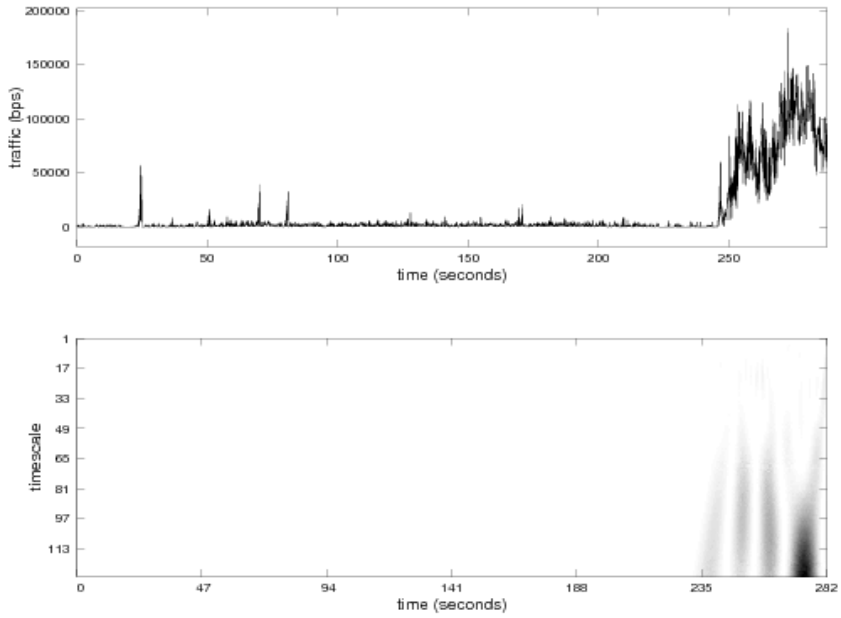

Figura 5.42 - Tráfego *downstream* SMTP por parte do servidor na direção B (bytes por segundo).

Relativamente à Figura 5.43, verifica-se uma situação análoga à Figura 5.39 no segmento de baixas frequências, com duas regiões diferenciadores ao nível da atividade do cliente (região A é relativa às situações em que o utilizador é pouco ativo e a região B é relativa às situações em que o utilizador tem uma atividade mais intensa). No segmento de médias frequências também se encontram duas regiões: a região C, que congrega eventos de média frequência com variação de energia relativamente intensa e a região D que envolve eventos com variação de energia mais reduzida, ou seja, com menor criação de sessões TCP pelo cliente com o servidor. A região C, como envolve várias sessões,

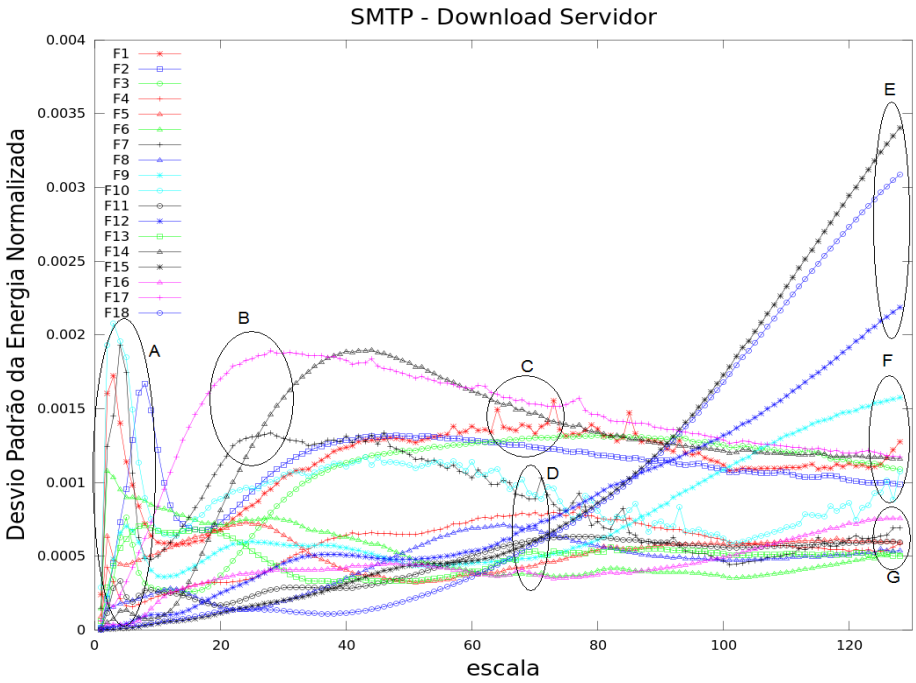

Figura 5.43 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* SMTP (do ponto de vista do servidor).

terá maior probabilidade de estar ligada à atividade de mais clientes, pelo menos comparativamente à região D. No que concerne ao segmento de altas frequências, encontram-se três regiões com características distintas. A região E engloba eventos com grande percentagem de componentes de alta frequência, relativos à transmissão de um grande volume de pacotes (neste caso, pode dever-se ao envio de um ou mais emails por parte dos clientes com ficheiros em anexo, o que pode explicar o perfil de tráfego nesta região). Por outro lado, a região G inclui tráfego com número reduzido de eventos, o que alude a uma atividade reduzida mais reduzida do(s) cliente(s), havendo apenas troca de pacotes relativos à monitorização e sincronização da ligação entre a interface da aplicação de email e o servidor remoto. Entre estas duas regiões situa-se a região F, que possui uma percentagem superior de componentes de alta frequência comparativamente com a região G, ou seja, com atividade mais intensa por parte dos clientes.

5.3 POP3

5.3.1 Cliente (*Downstream*)

Para a situação do tráfego *downstream* POP3 por parte do cliente foram encontrados três exemplos diferentes. No primeiro exemplo (Figura 5.44) temos tráfego intenso e pseudo periódico de pacotes a partir do segundo minuto da amostra, com muitos picos de tráfego de curta duração e grande amplitude. É possível observar que por volta dos 170 segundos o tráfego diminui abruptamente, ao qual se seguem novos picos de tráfego, o que indica a passagem de uma sessão TCP para a seguinte. Os picos de tráfego do início de cada sessão estão associados sobretudo a componentes de média e alta frequência, o que valida a alegação de surgirem devido a tráfego de sincronização de cada sessão.

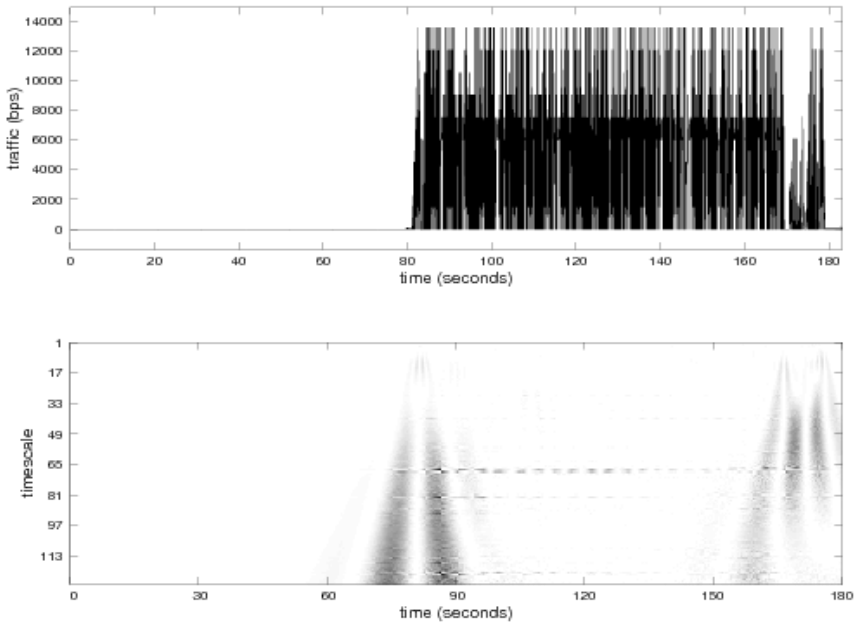

Figura 5.44 - Tráfego *downstream* POP3 por parte do cliente na direção A (bytes por segundo).

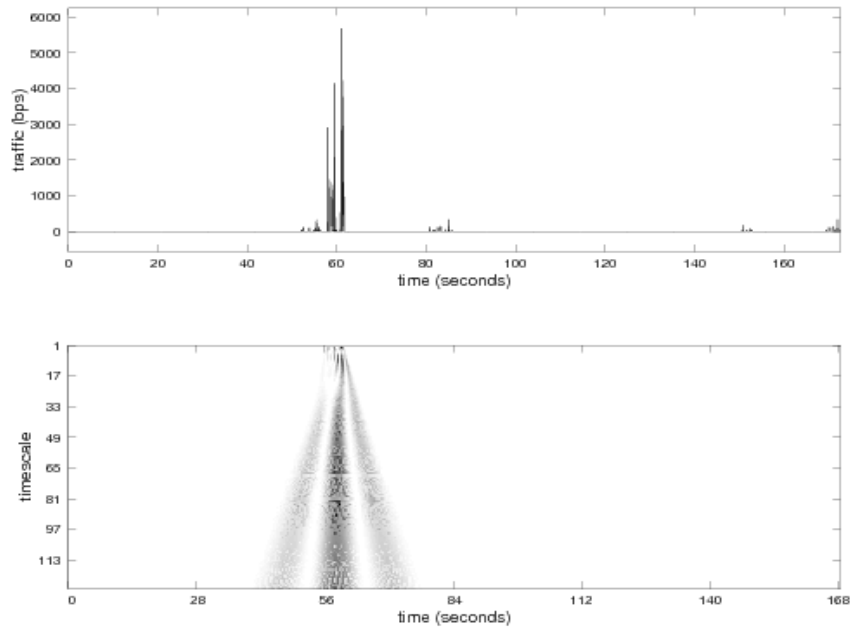

Figura 5.45 - Tráfego *downstream* POP3 por parte do cliente na direção A (bytes por segundo).

A presença de componentes de alta frequência relevantes indicia transferência de pacotes, ou seja, receção de um ou mais emails. O prolongamento no tempo do tráfego intensivo de pacotes pode dever-se à receção de emails de grande tamanho, fazendo com que a sua transmissão seja mais demorada. No que concerne o segundo exemplo (Figura 5.45), ocorre apenas um pico de tráfego no final do primeiro minuto da amostra, com duração moderada e amplitude relativamente grande. Estão associados a este pico de tráfego componentes de toda a gama de frequências, mas principalmente componentes de média e alta frequência. Pode então alegar-se que este pico de tráfego constitui a abertura da caixa de correio por parte do utilizador, sendo que este recebe um ou mais emails de pequeno tamanho.

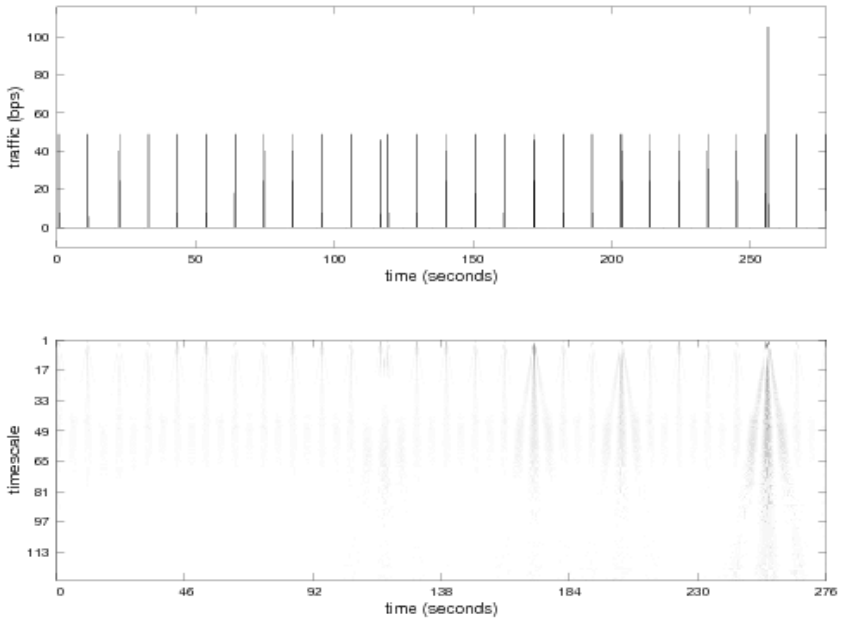

Figura 5.46 - Tráfego *downstream* POP3 por parte do cliente na direção B (bytes por segundo).

Já a Figura 5.46 constitui um caso em que existem vários picos de tráfego de curta duração e baixa amplitude, pelo que as componentes de frequência associadas aos mesmos são bastante fracas. O carácter periódico destes picos de tráfego e a sua amplitude significam que o cliente de email verifica periodicamente a existência de novos emails na caixa de entrada. Observa-se também a existência de um pico de tráfego de maior amplitude comparativamente aos restantes, o que alude ao descarregamento de algumas mensagens por parte do cliente de email.

Relativamente à Figura 5.47, a região A envolve fluxos de tráfego que geram eventos de muito baixa frequência, normalmente criados pela abertura da interface da aplicação de email e pela sincronização automática da ligação entre a interface da aplicação do lado do cliente e o servidor remoto. Já a região B contém fluxos de tráfego com percentagens relevantes de componentes de média frequência, o que indicia criação de algumas sessões TCP devido à consulta de emails ou outras funcionalidades da aplicação de email. Os fluxos de tráfego com maior periodicidade situam-se na região B, pois as interações POP3 são em maior número comparativamente aos fluxos em que o tráfego tem uma intensidade menor e que se encontram na região C. No segmento das altas frequências encontram-se três regiões. A região D envolve três fluxos (fluxos 1,3 e 10) com grandes percentagens de componentes de altas frequências, o que alude a um tráfego de pacotes *downstream* para o cliente bastante intenso, com possível receção de um ou mais emails; nesta região poderão existir fluxos com grande periodicidade, o que se traduz em maior tráfego de pacotes, comparativamente às regiões E e F. Por outro lado, a região F caracteriza-se por eventos com uma percentagem reduzida de componentes de alta frequência, o que indicia que nestes fluxos houve tráfego de pacotes *downstream* de baixa intensidade, e em que possivelmente os clientes de email verificam a caixa de email poucas vezes, o que se traduz num tráfego de pacotes menor (fluxos 5,8 e 9). Entre estas duas regiões situa-se a região E que apresenta uma taxa de chegada de pacotes considerável, mas inferior à taxa registada na região D.

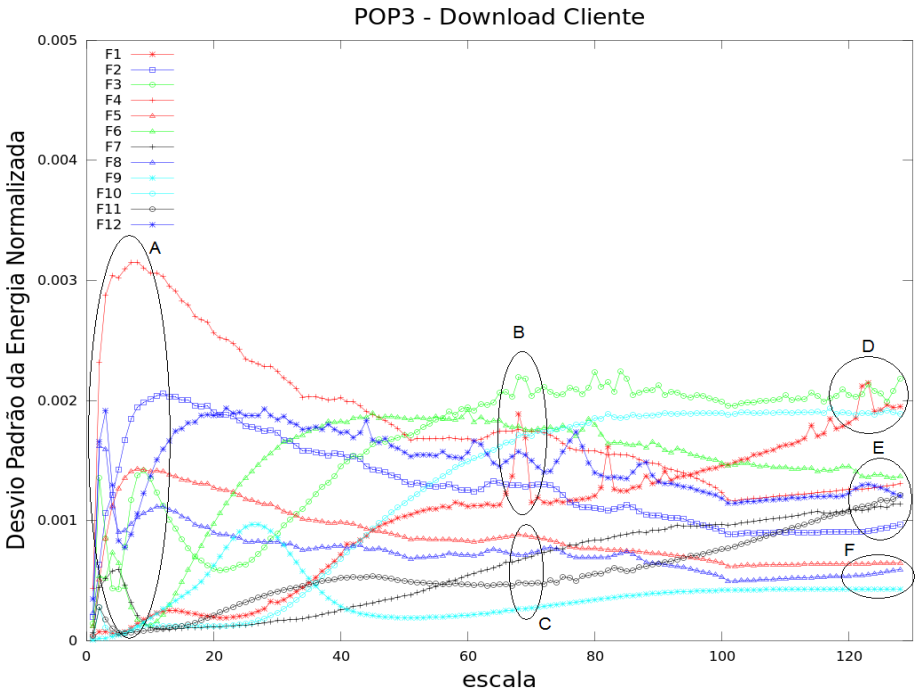

Figura 5.47 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* POP3 (do ponto de vista do cliente).

5.3.2 Servidor (*Upstream*)

Analisando a Figura 5.48, verifica-se que o tráfego tem um cariz não periódico e os picos de tráfego têm amplitude idêntica e curta duração. Estes picos estão associados sobretudo a componentes de média frequência e por vezes a componentes de alta frequência. Logo, tendo em conta o tamanho destes pacotes, eles fazem parte da monitorização das ligações estabelecidas entre o utilizador e o servidor e por vezes são também enviados pequenos emails; nos picos de tráfego de maior duração existem mais pedidos de clientes e como tal o servidor envia mais pacotes de resposta.

No que diz respeito à Figura 5.49, existe um pico de tráfego bem definido de moderada duração e com grande amplitude, associado a fortes componentes de média e alta frequência, o que indicia transferência de ficheiros do servidor para o utilizador.

Na Figura 5.50, o tráfego de pacotes é não periódico e ocorrem vários picos de tráfego sobretudo com média amplitude e curta duração, associados sobretudo a pequenos componentes de média e alta frequência no escalograma. Neste caso, existem vários clientes a efetuarem pedidos ao servidor, daí o tráfego ser tão intenso e quase não ter falhas.

Analisando a Figura 5.51, encontram-se duas regiões demarcadas no segmento de baixas frequências. A região A abrange eventos na zona das frequências muito baixas, gerados pela resposta do servidor à abertura por parte do cliente da interface da aplicação de email e consequente sincronização automática da ligação entre o cliente e o servidor remoto para a verificação da caixa de correio. A região B contém eventos de baixa frequência originados pela resposta do servidor a cliques do cliente quando este verifica a sua aplicação de email e pela atualização periódica da caixa de correio da mesma aplicação. As regiões C e D situam-se no segmento das médias frequências e

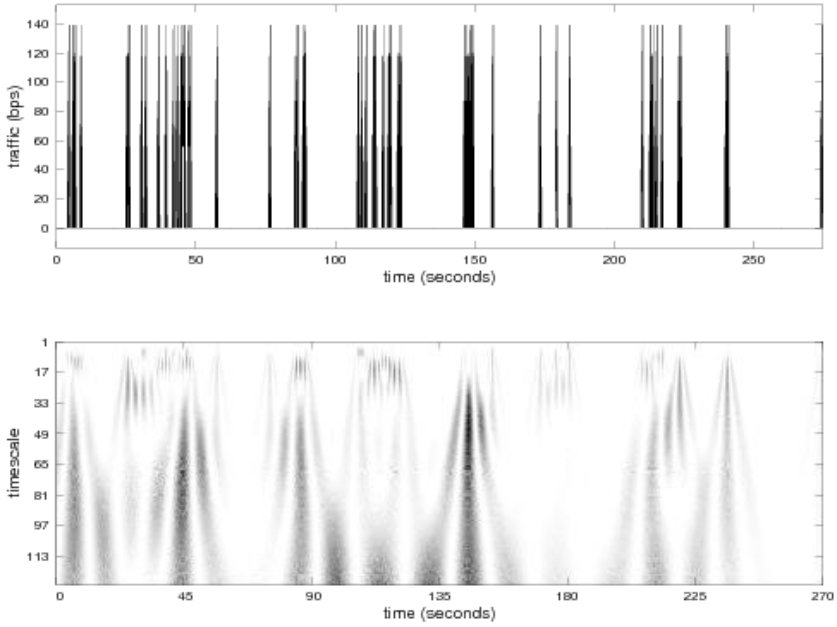

Figura 5.48 - Tráfego *upstream* POP3 por parte do servidor na direção B (bytes por segundo).

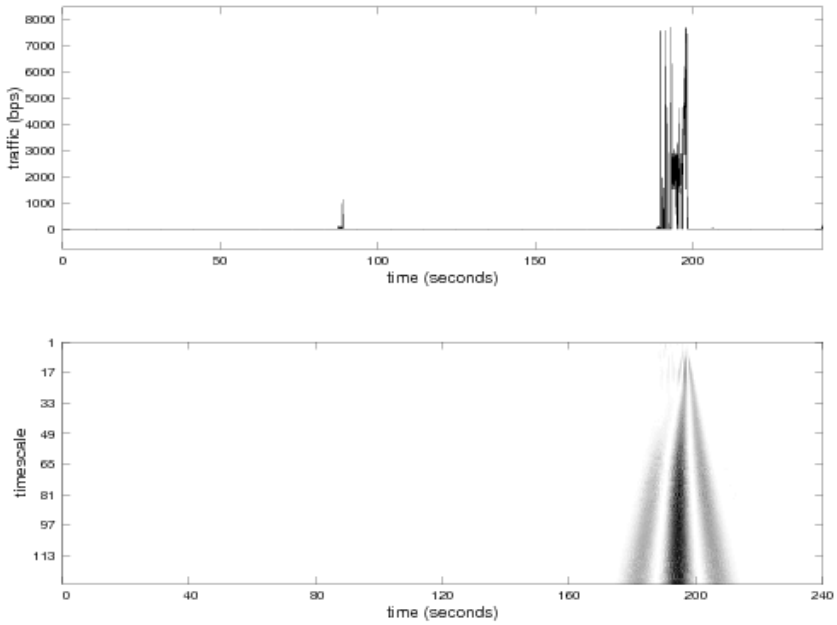

Figura 5.49 - Tráfego *upstream* POP3 por parte do servidor na direção A (bytes por segundo).

diferenciam os eventos em que o servidor é muito solicitado para a abertura de várias sessões TCP e visualização de conteúdo na interface da aplicação de email (região C) dos eventos em que o servidor é pouco requerido para a criação de sessões TCP e portanto a atividade do cliente é mais branda (região D). Tal como na Figura 5.47, existem três regiões no segmento de altas frequências, embora neste caso em análise a região E contenha apenas um fluxo (fluxo 4) com grande percentagem de componentes de alta frequência (o que indicia que o servidor tenha enviado um ou mais emails, tendo como destinatários diferentes clientes). A região F possui eventos com percentagem razoável de componentes de alta frequência, apesar de ter uma taxa de chegada de pacotes inferior comparativamente à região E. Finalmente, a região G (fluxos 1 e 7)

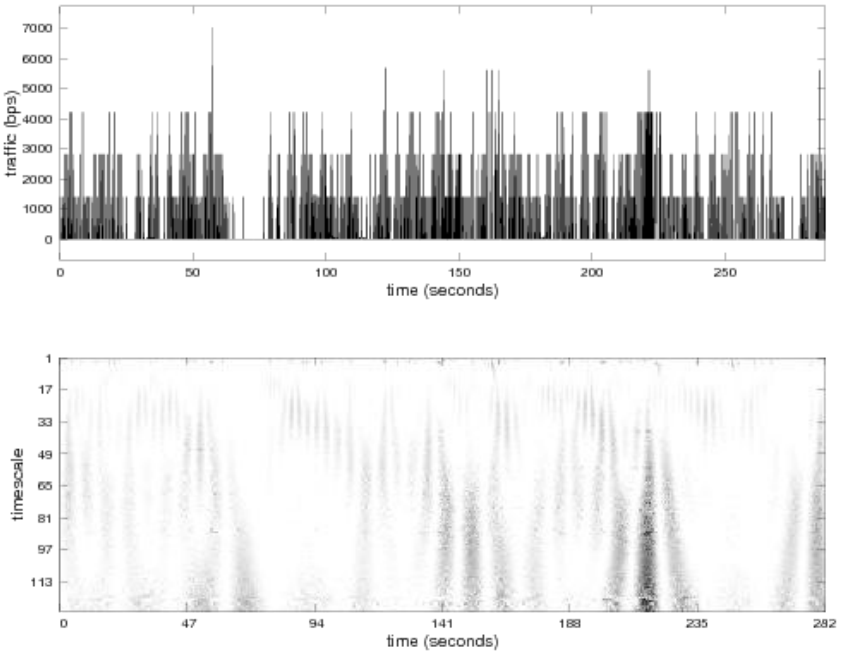

Figura 5.50 - Tráfego *upstream* POP3 por parte do servidor na direção B (bytes por segundo).

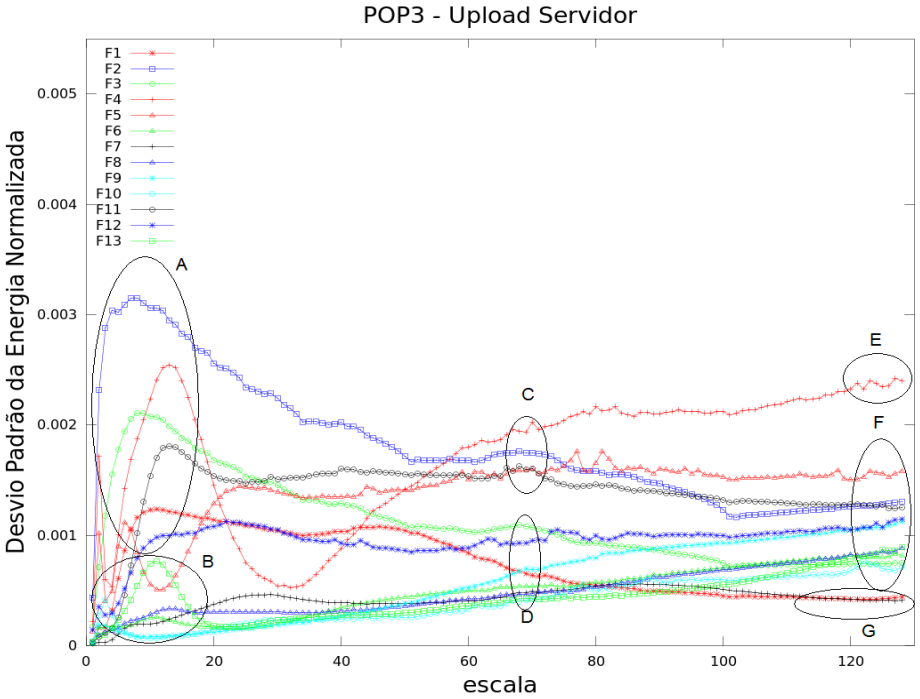

Figura 5.51 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* POP3 (do ponto de vista do servidor).

agrega eventos com uma percentagem diminuta de componentes de alta frequência, logo nestes casos o servidor apenas troca pacotes de controlo e sincronização da sua ligação com a interface da aplicação de email do utilizador e o número de clientes envolvidos é menor comparativamente aos envolvidos nas regiões E e F.

5.3.3 Cliente (*Upstream*)

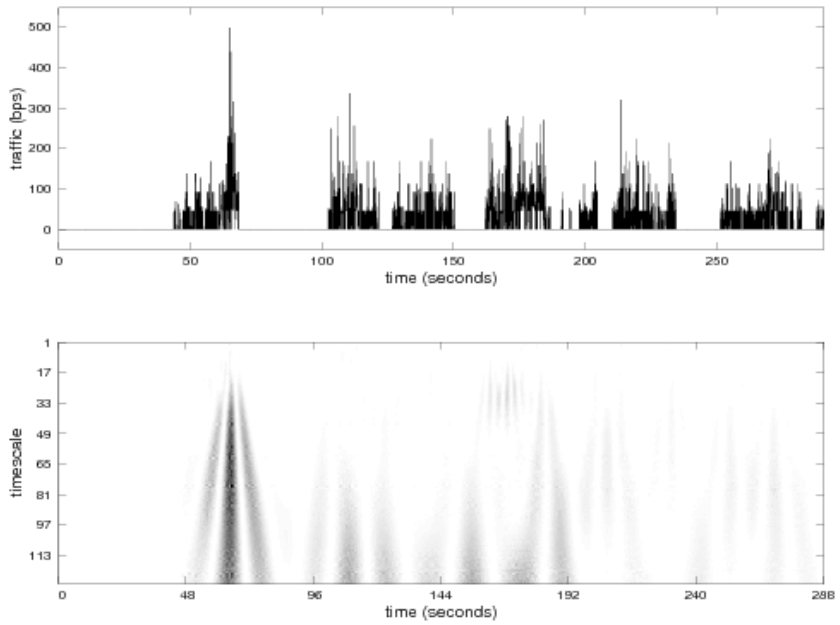

Figura 5.52 - Tráfego *upstream* POP3 por parte do cliente na direção A (bytes por segundo).

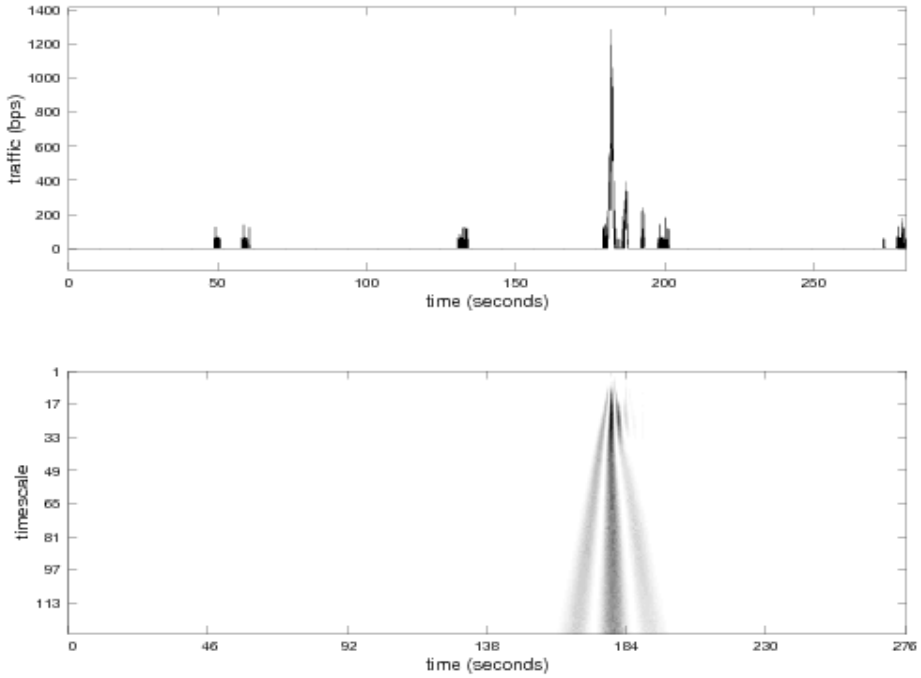

Figura 5.53 - Tráfego *upstream* POP3 por parte do cliente na direção B (bytes por segundo).

Para a situação do de tráfego *upstream* POP3 por parte do cliente foram encontrados três exemplos diferentes. No primeiro exemplo (Figura 5.52) verifica-se que os pacotes surgem em agregados, com duração de dezenas de segundos cada. Existem poucos picos de tráfego, sendo que quase todos eles são de curta duração e baixa amplitude. Apenas um pico de tráfego tem componentes de média e alta frequência associados, que poderão

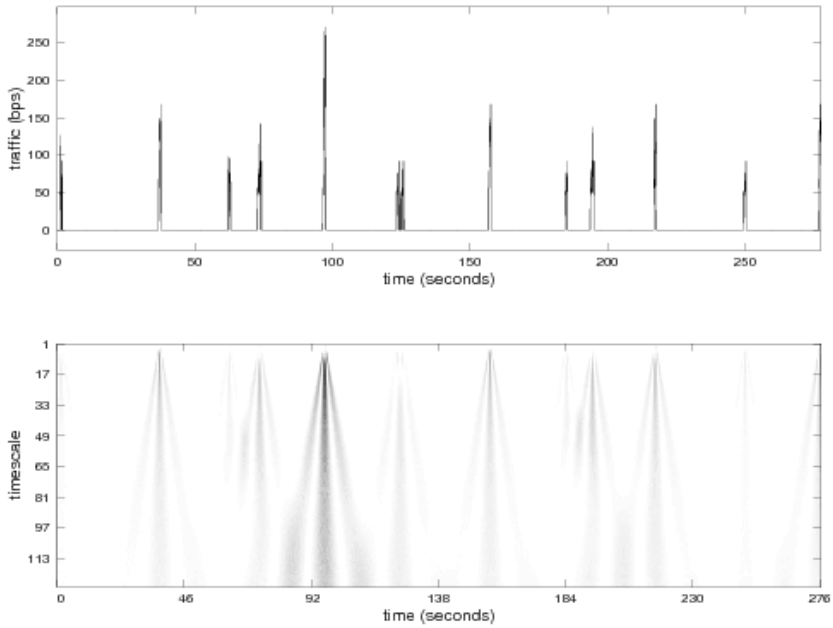

Figura 5.54 -Tráfego *upstream* POP3 por parte do cliente na direção A (bytes por segundo).

estar relacionados com a sincronização da ligação entre o cliente de email do lado do cliente e o servidor tendo em conta a sessão iniciada pelo primeiro ou o envio de pacotes de reconhecimento pela chegada de algum email. No que concerne à Figura 5.53, regista-se apenas um pico de tráfego com moderada duração, mas de considerável amplitude, responsável pela criação de componentes de média e alta frequência no escalograma, que poderá advir do envio de pacotes de resposta à chegada de um email à caixa de um cliente. Tendo em conta os poucos pacotes enviados, pode assumir-se que o número de clientes ativos é diminuto.

No último exemplo (Figura 5.54) observa-se a ocorrência de picos de tráfego pseudo periódicos de curta duração, responsáveis pela aparição de vários componentes de baixa e média frequência no escalograma, o que indicia que este caso reporta-se a cliques do cliente ao aceder à sua caixa de correio ou então pacotes enviados pelo cliente de email para sincronização automática da ligação e verificação da chegada de novos emails.

No que concerne à Figura 5.55, esta apresenta uma disposição das regiões ao longo do gráfico semelhante à Figura 5.51: duas regiões tanto no segmento de baixas frequências como no segmento de médias frequências e três regiões no segmento de altas frequências. Contudo, é importante ressalvar que a região B abrange eventos de frequência muito baixa com variação de energia diminuta, o que indica atividade rara do cliente e o tráfego presente nesta região deriva sobretudo das sincronizações automáticas entre a interface da aplicação de email do cliente e o servidor remoto. Já a região A engloba eventos de baixa frequência mas com variação de energia considerável, o que pressupõe uma maior atividade do utilizador. As regiões C e D situam-se no segmento das médias frequências e diferenciam os eventos em que o cliente solicita ao servidor a abertura de várias sessões TCP e visualização de conteúdo na interface da aplicação de email (região C) dos eventos em que o cliente tenta criar poucas sessões TCP e portanto o cliente não está tão ativo (região D). No segmento das altas frequências, a região E (fluxos 2,11 e 17) agrega os eventos com percentagens elevadas de componentes de alta frequência possivelmente relacionados com os pacotes enviados

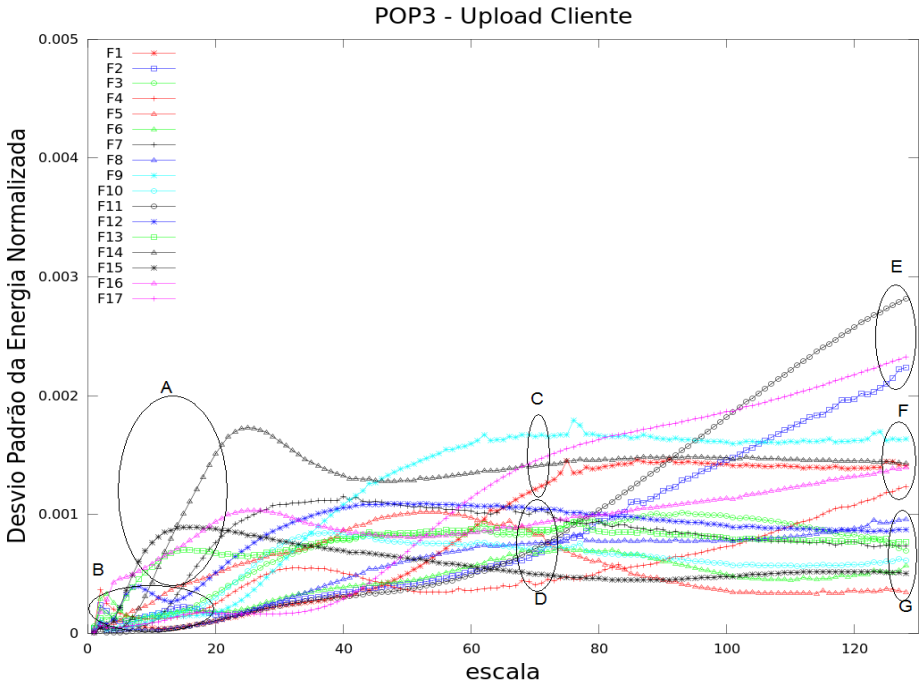

Figura 5.55 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* POP3 (do ponto de vista do cliente).

pelo cliente para o controlo da chegada de um ou mais emails. A região F contém eventos com percentagens consideráveis de componentes de alta frequência, mas com taxa de transmissão de pacotes inferior à região E. Finalmente, a região G compreende eventos com variação de energia reduzida, o que indicia uma menor transmissão de pacotes por parte do cliente.

5.3.4 Servidor (*Downstream*)

Analisando a Figura 5.56, observa-se a ocorrência de picos de tráfego com duração de alguns segundos, durante os quais as componentes de média e alta frequência têm relativa intensidade. Durante este período, o tráfego de pacotes do utilizador para o servidor é intenso e refere-se sobretudo ao estabelecimento de ligações e ao envio de pacotes de reconhecimento (*acknowledge*) em relação ao conteúdo recebido na caixa de correio do utilizador. A mesma análise pode aplicar-se à Figura 5.57, embora neste caso só é visível um pico de tráfego. Na Figura 5.56, o facto de haver falhas na transmissão do tráfego e picos de tráfego bastante espaçados no tempo indica que há poucos clientes a enviar pedidos para o servidor. Na Figura 5.57, o tráfego é constante e contínuo, portanto resulta da soma de vários fluxos provenientes de diferentes clientes.

Relativamente à Figura 5.58, o tráfego não tem a regularidade dos dois casos anteriores, o que significa que o tráfego está bastante espaçado no tempo e os picos de tráfego estão claramente identificados. A amplitude dos picos de tráfego presentes é normalmente baixa, apesar das componentes de média e alta frequência do escalograma serem bastante intensas nos picos de tráfego do último minuto da amostra, a que não é alheio o facto de serem de maior duração comparativamente aos picos de tráfego dos

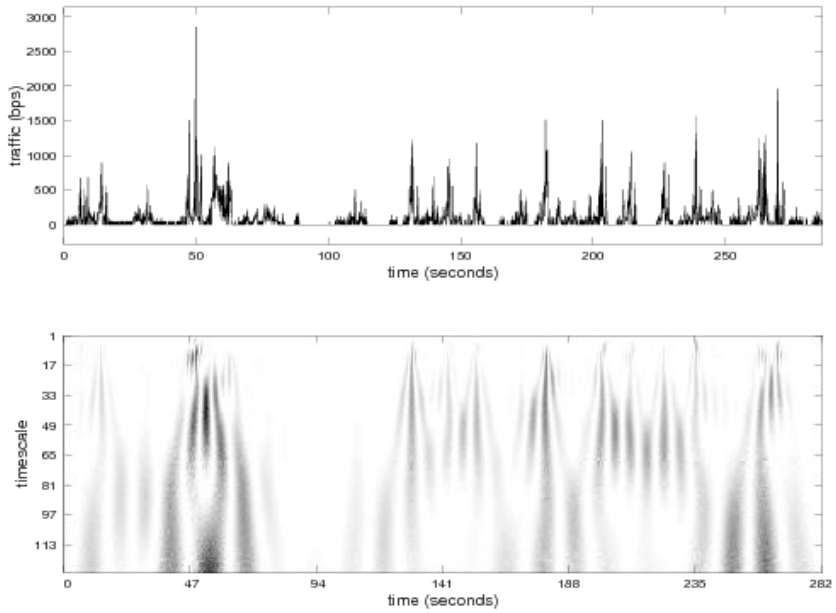

Figura 5.56 - Tráfego *downstream* POP3 por parte do servidor na direção B (bytes por segundo).

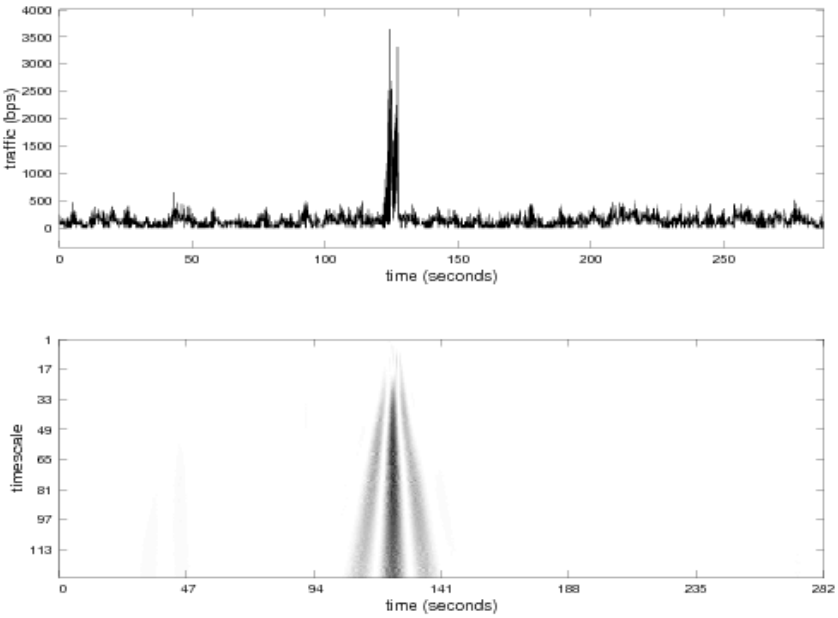

Figura 5.57 - Tráfego *downstream* POP3 por parte do servidor na direção B (bytes por segundo).

minutos anteriores. Tendo em conta os grandes intervalos de tempo em que não há tráfego, pode assumir-se que neste cenário existem poucos clientes a enviar informação ou então os mesmos estão muito tempo inativos.

No que diz respeito à Figura 5.59, a análise das regiões A e B no segmento de baixas frequências faz-se de forma semelhante à Figura 5.55, com a diferença de neste caso o fluxo de tráfego ser *downstream* em direção ao porto do servidor. A região C engloba todos os eventos de média frequência que neste cenário apresentam variação de energia com amplitude pequena ou moderada, o que indica que enquanto nalguns

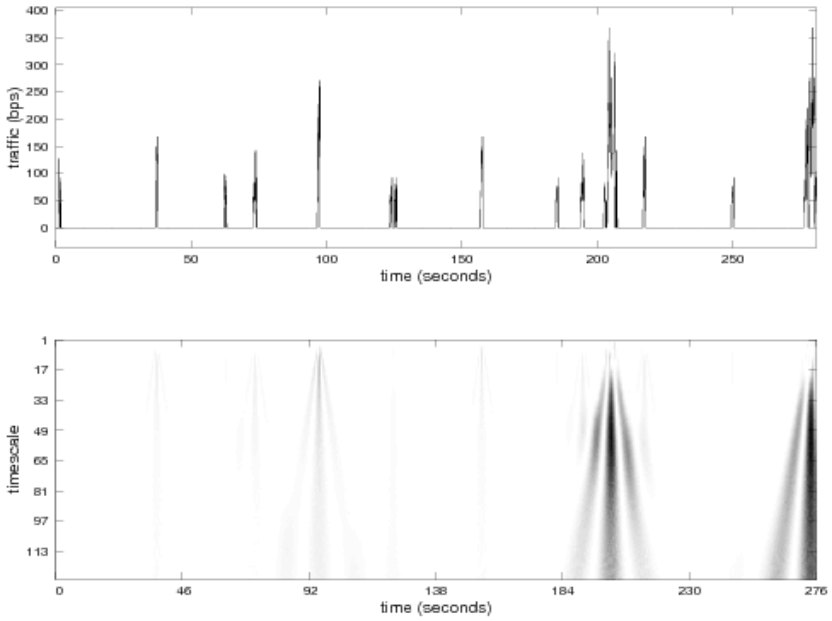

Figura 5.58 - Tráfego *downstream* POP3 por parte do servidor na direção B (bytes por segundo).

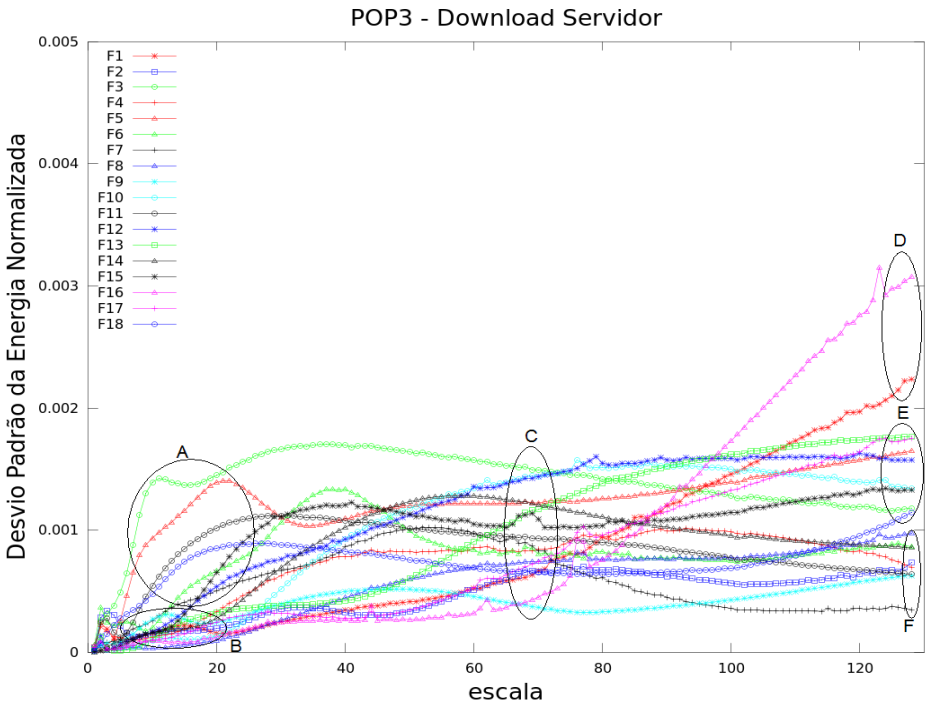

Figura 5.59 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* POP3 (do ponto de vista do servidor).

fluxos de tráfego houve bastante interações TCP e POP3, noutros casos estas interações são menos frequentes e limitam-se ao controlo e sincronização da ligação entre cliente e servidor. Relativamente ao segmento de altas frequências, na região F encontram-se eventos com variação de energia de pequena amplitude, o que alude a um tráfego reduzido de pacotes para o servidor, o que significa presença de poucos clientes e com pouca atividade por parte dos mesmos. A região E já apresenta uma taxa considerável de

transmissão de pacotes (a variação de energia destes eventos já é de maior amplitude) enquanto a região D (fluxos 1 e 16) abrange eventos com variação de energia com grande amplitude, o que implica um grande volume de pacotes transmitidos para o servidor, ou seja possivelmente um ou mais clientes receberam emails e estes são os pacotes de reconhecimento e de controlo da operação de receção que o(s) cliente(s) envia(m) para o servidor para certificar que essa mesma operação decorreu normalmente.

5.4 IMAP

5.4.1 Cliente (*Downstream*)

Foram encontrados três exemplos diferentes para a situação do tráfego *downstream* IMAP por parte do cliente. No primeiro exemplo (Figura 5.60), existem vários picos de tráfego concentrados sobretudo no último minuto da amostra. Estão associados a componentes de toda a gama de frequências, com relevância para os componentes de alta frequência; logo estes picos de tráfego estão associados à receção de um ou mais emails de tamanho pequeno.

No segundo exemplo (Figura 5.61), ocorrem múltiplos picos de tráfego de curta duração e com amplitudes moderadas a altas. Os componentes de baixa frequência estão espalhados ao longo do escalograma e os componentes de alta frequência têm uma presença bastante pequena, o que indicia que se o cliente receber emails, estes serão de pequeno tamanho. Finalmente na Figura 5.62 pode observar-se que os picos de tráfego são periódicos e de curta duração. A disposição das componentes de frequência no escalograma deixa a perceber que os picos de tráfego de menor amplitude são originados pelo cliente de email ao verificar periodicamente se chegaram novos emails à caixa de correio, enquanto os picos de tráfego de maior amplitude são originados pela receção de emails pequenos.

Analisando a Figura 5.63, observa-se a presença de eventos na zona das frequências muito baixas (região A), gerados por acontecimentos raros como o download da interface da aplicação de email do cliente e posteriores sincronizações automáticas da ligação entre servidor e cliente. Na região B encontram-se eventos de baixa frequência com variação de energia pequena, gerados por cliques do cliente ao aceder à sua aplicação de email. O segmento de médias frequências está dividido em duas regiões, sendo que uma delas abrange apenas dois fluxos (fluxos 1 e 4) com variação de energia bastante elevada; ambos os fluxos são responsáveis por eventos em que existem bastante interação TCP, associada à consulta de informação por parte do cliente e à verificação periódica do cliente de email da chegada de novos emails. Já a região D abrange eventos com menor variação de energia, onde a atividade do cliente é também menos intensa e a periodicidade da atividade do cliente de email também é inferior. O segmento de altas frequências está dividido em duas regiões: a região E engloba eventos com percentagem elevada de componentes de alta frequência, ou seja, existe transmissão *downstream* de pacotes em grande quantidade (associados à receção de emails).

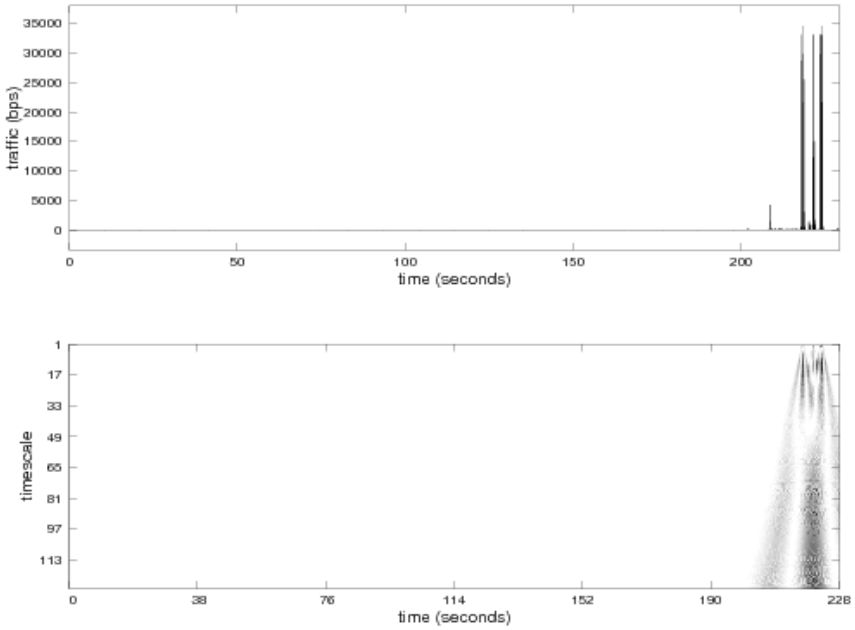

Figura 5.60 - Tráfego *downstream* IMAP por parte do cliente na direção A (bytes por segundo).

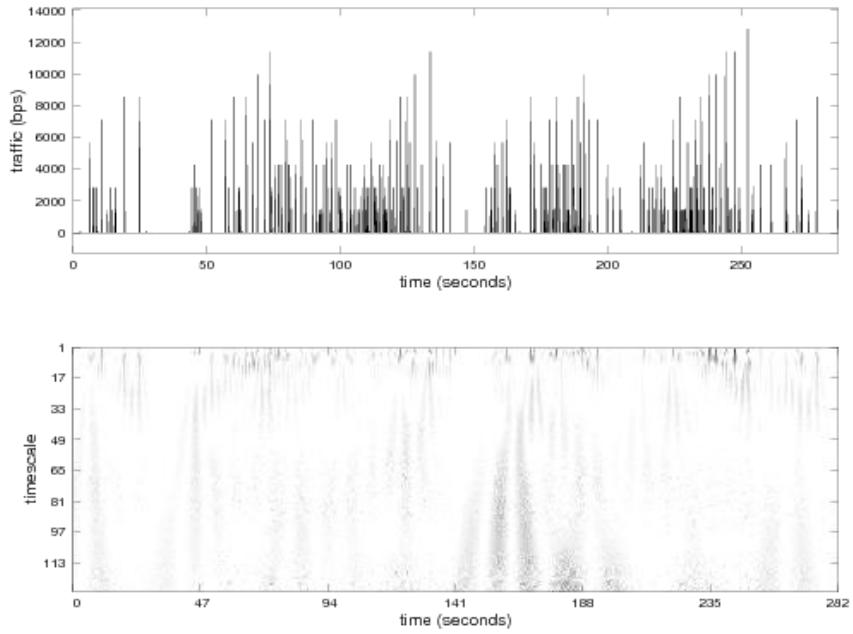

Figura 5.61 - Tráfego *downstream* IMAP por parte do cliente na direção B (bytes por segundo).

Quanto à região F, esta abrange eventos com uma percentagem reduzida de componentes de alta frequência, o que aluda a transmissão reduzida de pacotes *downstream*, logo nesta região se houver receção de emails serão em menor quantidade e com tamanho menor comparativamente à região E.

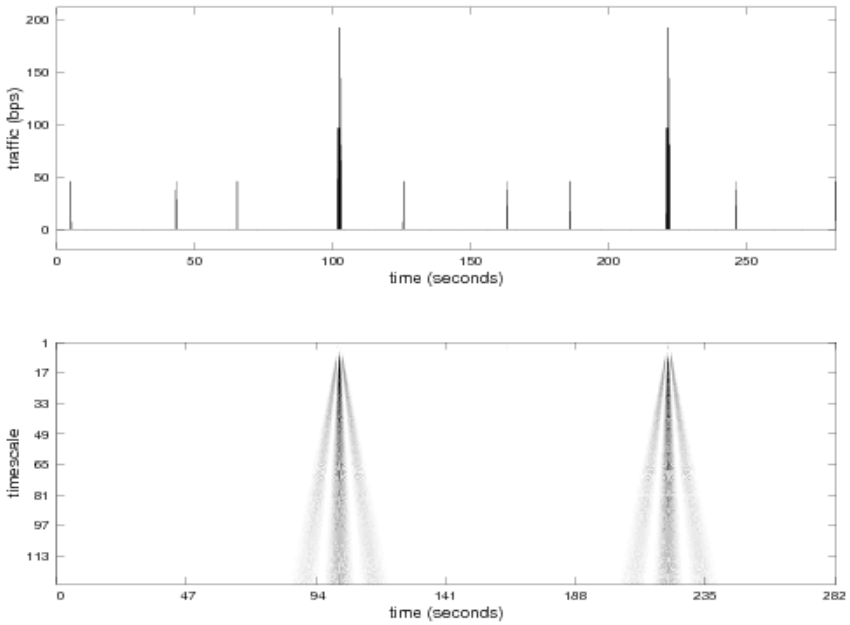

Figura 5.62 - Tráfego *downstream* IMAP por parte do cliente na direção B (bytes por segundo).

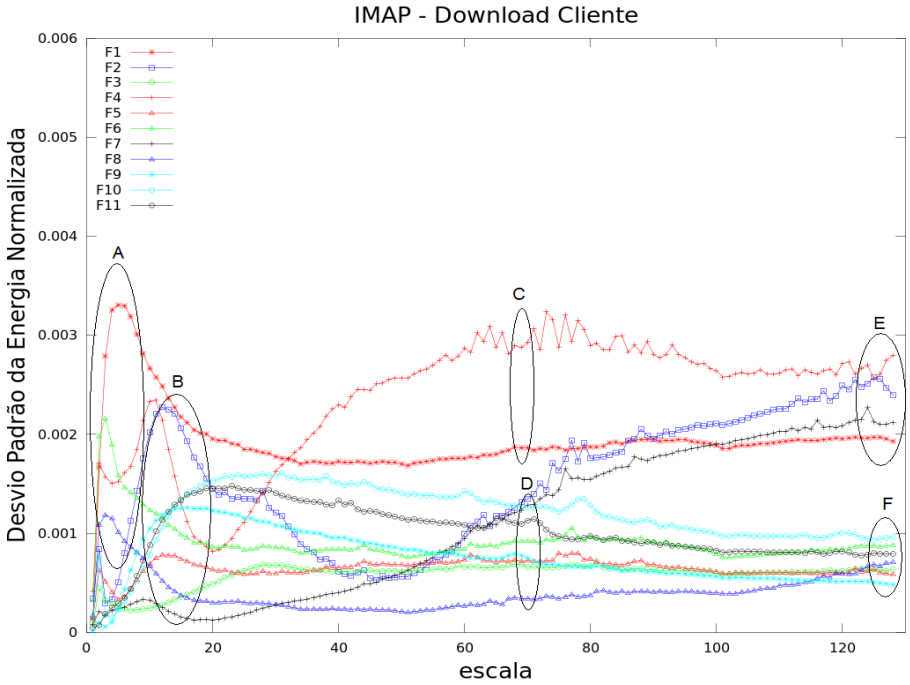

Figura 5.63 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* IMAP (do ponto de vista do cliente).

5.4.2 Servidor (*Upstream*)

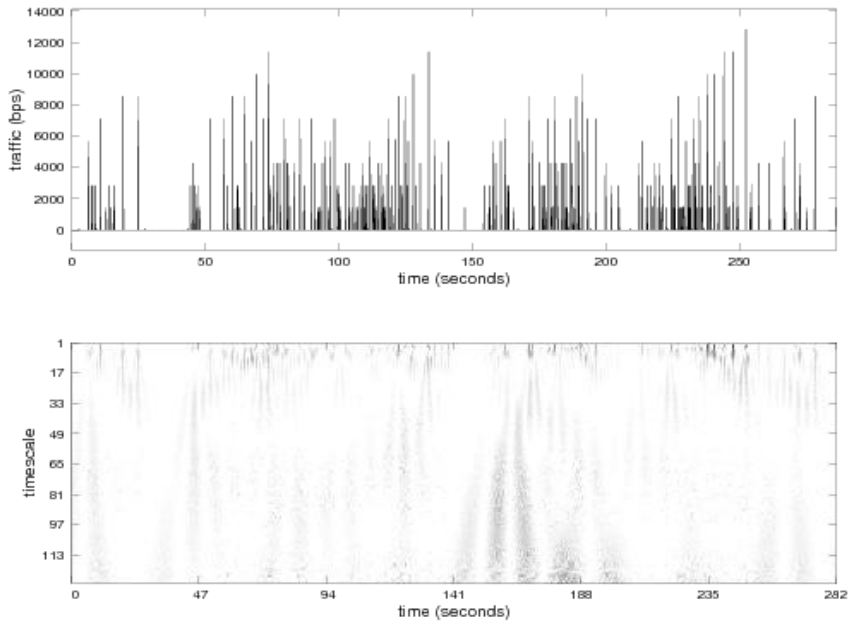

Figura 5.64 - Tráfego *upstream* IMAP por parte do servidor na direção B (bytes por segundo).

A Figura 5.64 e a Figura 5.65 são dois exemplos de tráfego *upstream* IMAP por parte do servidor. No caso da Figura 5.64, verifica-se que o tráfego é não periódico, com vários picos de tráfego de amplitude moderada e alta. Praticamente todos estes picos de tráfego estão associados a componentes de baixa frequência no escalograma, o que indica que funcionam como resposta a pedidos do utilizador. A existência de vários picos de tráfego de diferentes amplitudes e o facto de o tráfego não ser contínuo em todo o intervalo temporal permitem assumir que o servidor atende pedidos de poucos clientes e estes não estão sempre ativos.

O caso da Figura 5.65 é diferente, pois existe tráfego pseudo periódico durante períodos de tempo extremamente curtos e os picos de tráfego têm amplitudes mais baixas comparativamente ao caso anterior. Contudo, estes picos de tráfego têm componentes de praticamente toda a gama de frequências, o que indica que funcionam como resposta a pedidos do utilizador, abrindo sessões entre servidor e o cliente e transferindo conteúdo. Tendo em conta a pouca quantidade de tráfego enviado pelo servidor e os grandes intervalos de tempo em que não se regista qualquer tipo de tráfego, pode concluir-se que o servidor envia emails curtos e para poucos clientes.

A Figura 5.66 tem a delimitação de regiões em cada segmento de frequência de forma similar à efetuada na Figura 5.63, com a ressalva que neste caso o tráfego é observado do ponto de vista do servidor, ou seja, este tráfego é gerado em resposta a solicitações do cliente. É importante realçar que contrariamente ao sucedido na situação do tráfego downstream do ponto de vista do cliente, a região C abrange mais fluxos de tráfego; portanto, é possível assumir que no cenário da Figura 5.66 as regiões C e E estão associadas a fluxos originados por servidores que enviam vários emails e trocam informação com vários clientes de forma mais intensa (e por vezes com um certa periodicidade) comparativamente aos fluxos das regiões D e F, em que os pacotes transmitidos e as sessões criadas são em menor número.

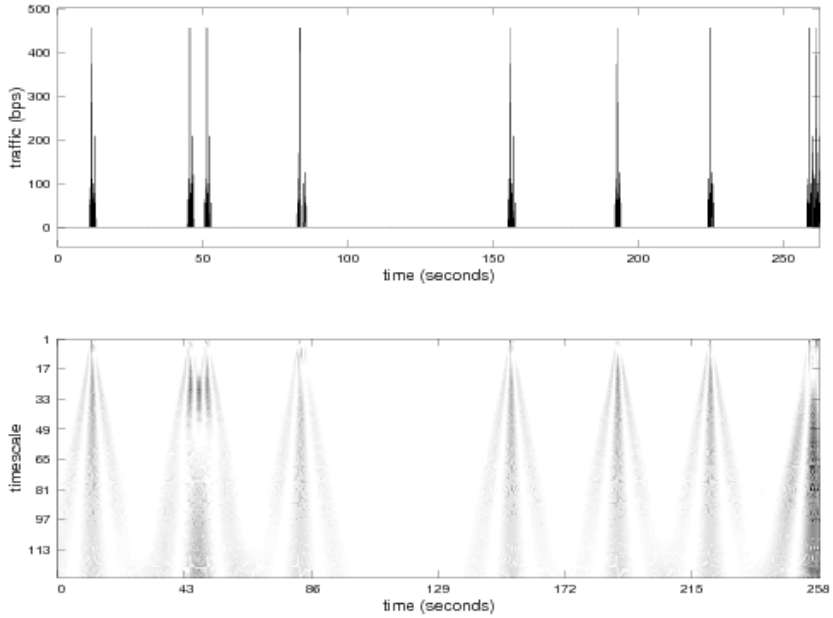

Figura 5.65 - Tráfego *upstream* IMAP por parte do servidor na direção A (bytes por segundo).

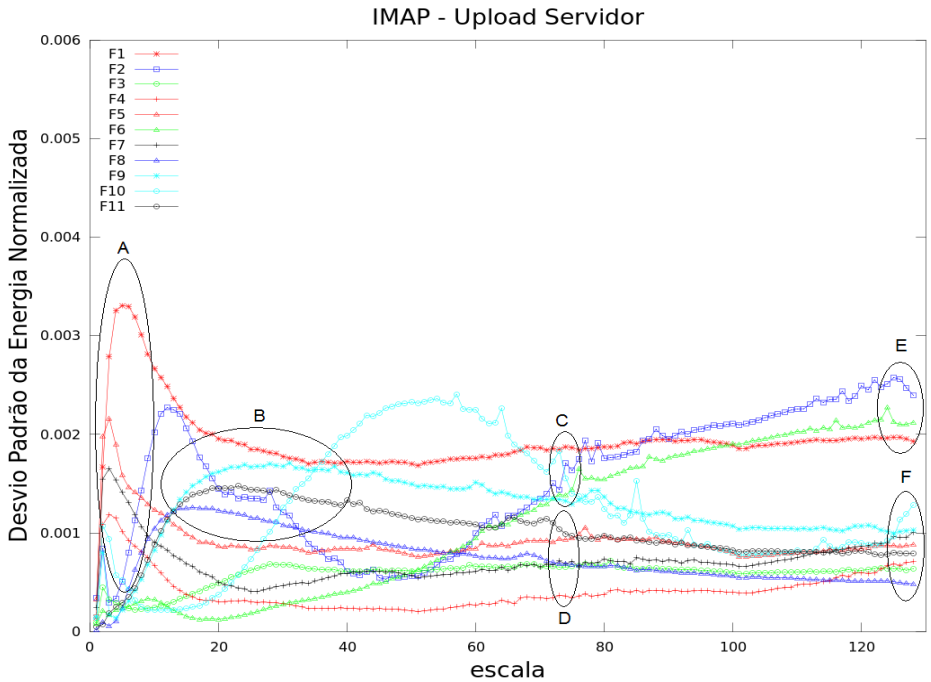

Figura 5.66 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* IMAP (do ponto de vista do servidor).

5.4.3 Cliente (*Upstream*)

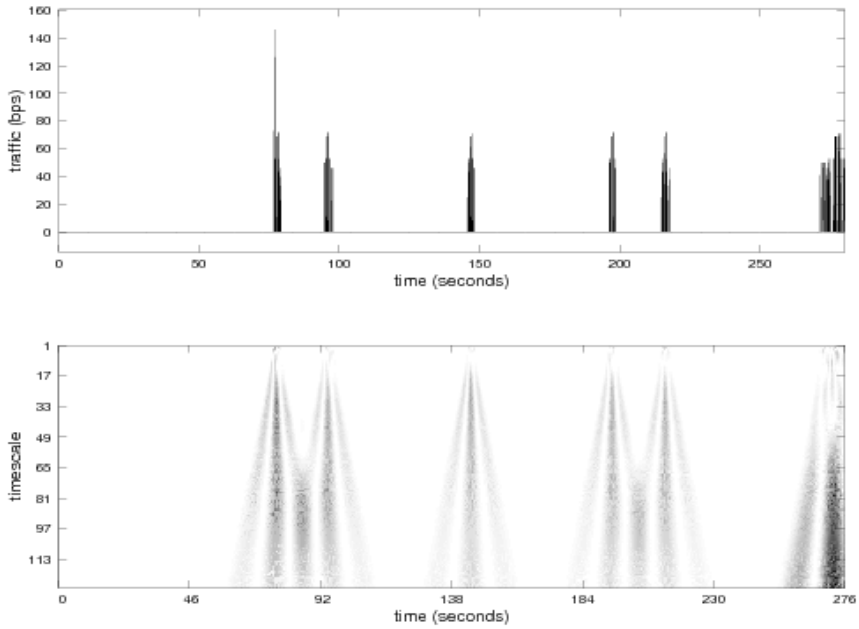

Figura 5.67 - Tráfego *upstream* IMAP por parte do cliente na direção A (bytes por segundo).

Analisa-se agora a situação de tráfego *upstream* IMAP por parte do cliente. A Figura 5.67 apresenta um perfil semelhante à Figura 5.62 em termos da periodicidade do tráfego. Contudo, neste caso trata-se de upload de tráfego e os picos de tráfego têm uma duração maior em comparação, o que se reflete no escalograma. As componentes de média e alta frequência têm relativa intensidade, mas tendo em conta o tamanho dos pacotes em causa este tráfego é sobretudo de monitorização da ligação entre o cliente e o servidor, embora o pico de tráfego de maior amplitude possa corresponder ao envio de pacotes de reconhecimento da receção de um email.

No caso da Figura 5.68, o tráfego *upstream* é mais intenso comparativamente à Figura 5.67, embora haja períodos sem qualquer evento. Há múltiplos picos de tráfego, com amplitudes a variar entre as pequenas e as médias. É possível observarem-se alguns componentes de média e alta frequência, mas de pequena intensidade, pois este tráfego é sobretudo de monitorização da ligação entre o cliente e o servidor.

Analisando a Figura 5.69, verifica-se que a região A engloba eventos de muito baixa frequência, normalmente gerados pela abertura da interface de aplicação do email por parte do cliente e pela sincronização automática da ligação entre cliente e servidor remoto. A região B engloba eventos de baixa frequência com variação de energia moderada, gerados por cliques do utilizador ao verificar a chegada de emails à sua caixa de correio eletrónico. No segmento de médias frequências encontram-se duas regiões, que diferenciam os fluxos de tráfego conforme tenham uma grande (região C) ou pequena variação de energia (região D). No segmento de altas frequências, três fluxos destacam-se por terem uma grande taxa de transmissão de pacotes enviados pelo cliente (fluxos 6,7 e 12), o que indicia que poderão ser pacotes de reconhecimento e controlo da receção de vários emails por parte do cliente. Na região G situam-se os fluxos de tráfego em que o utilizador tem uma atividade mais reduzida e a região F engloba os fluxos de tráfego que embora não sejam responsáveis por um grande volume de tráfego diferenciam-se por ter uma percentagem relevante de componentes de alta frequência.

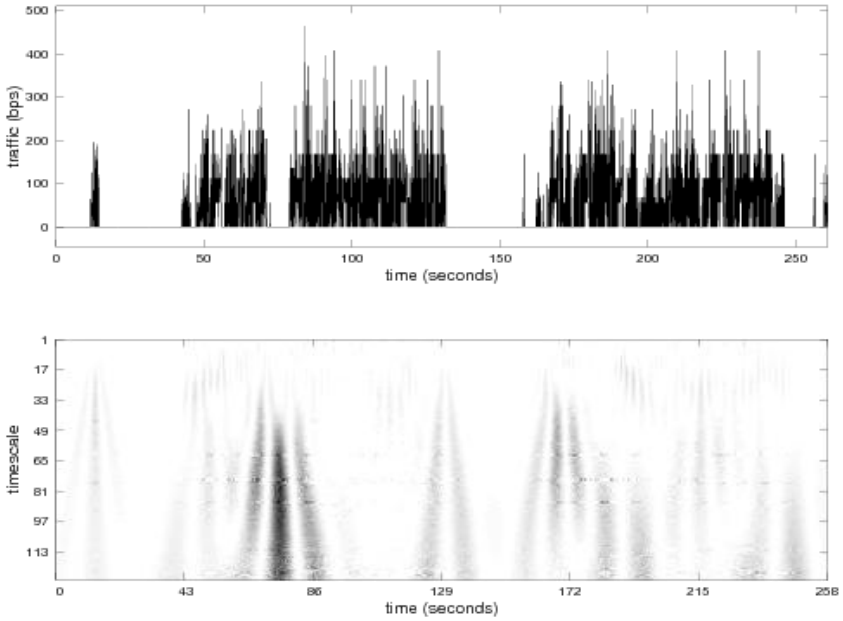

Figura 5.68 - Tráfego *upstream* IMAP por parte do cliente na direção B (bytes por segundo).

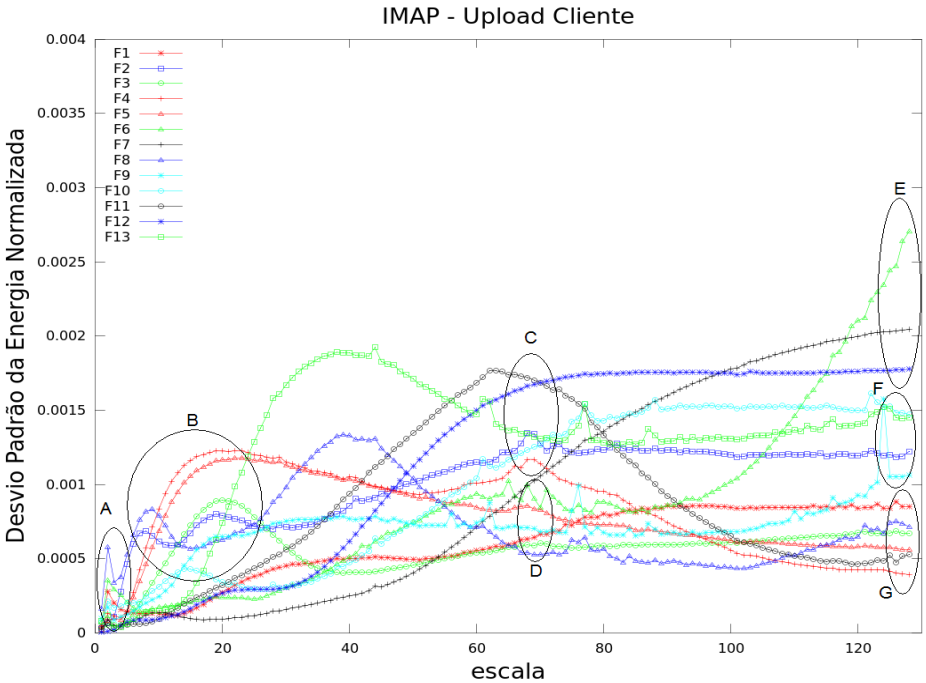

Figura 5.69 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* IMAP (do ponto de vista do cliente).

5.4.4 Servidor (*Downstream*)

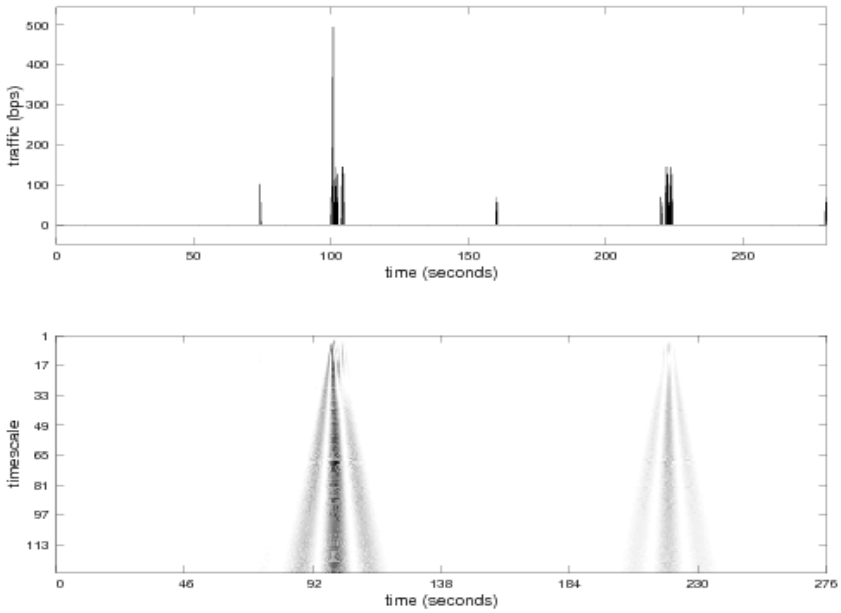

Figura 5.70 - Tráfego *downstream* IMAP por parte do servidor na direção A (bytes por segundo).

As figuras seguintes (Figura 5.70, Figura 5.71 e Figura 5.72) constituem exemplos de tráfego *downstream* IMAP por parte do servidor: um pico de tráfego em toda a amostra (Figura 5.70), tráfego intenso (Figura 5.71) e tráfego ocasional (Figura 5.72). Um ponto em comum entre estes três casos é o facto dos picos de tráfego possuírem uma amplitude moderada (apesar dos picos de tráfego da Figura 5.71 terem menor duração) e estarem associados a componentes de toda a gama de frequências. Também se pode dizer que os poucos clientes que enviam informação na Figura 5.70 e na Figura 5.72 encontram-se durante maior parte do tempo inativos, o que origina o tráfego ocasional presente nestas figuras. Na outra figura, os clientes são mais ativos, logo não existem tantas falhas no tráfego. Os escalogramas destes casos apresentam características similares em relação ao escalogramas obtidos para o download de tráfego POP3 por parte do servidor (Figura 5.58).

No que diz respeito à Figura 5.73, verifica-se que a região A abrange eventos com variação de energia moderada, o que aponta para alguma atividade do cliente ao aceder à interface da aplicação de email. A região B engloba eventos de baixa frequência com variação de energia diminuta, criados por pacotes de controlo enviados pelo cliente ou por atualizações periódicas da aplicação de email para verificação de chegada de novos emails. Como sucedido em cenários anteriores, o segmento de médias frequências está dividido em duas regiões, tendo em conta a variação de energia dos diferentes fluxos de tráfego. Na região C encontram-se os fluxos com mais interações IMAP, portanto com tráfego mais intenso e com maior periodicidade. No segmento de altas frequências verifica-se que um fluxo (fluxo 5) apresenta na região E uma grande percentagem de componentes de alta frequência, devido à transmissão de um elevado número de pacotes para o servidor, possivelmente pacotes de controlo e reconhecimento da receção de emails. A região F engloba eventos de alta frequência com percentagem considerável de componentes de alta frequência, embora menor comparativamente à região E. Finalmente, a região G abrange eventos de alta frequência com variação de energia

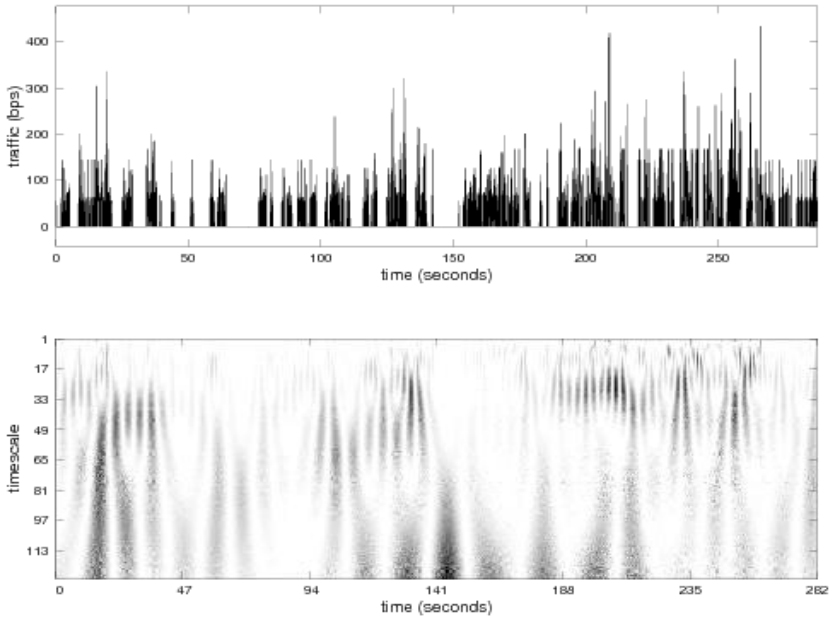

Figura 5.71 - Tráfego *downstream* IMAP por parte do servidor na direção B (bytes por segundo).

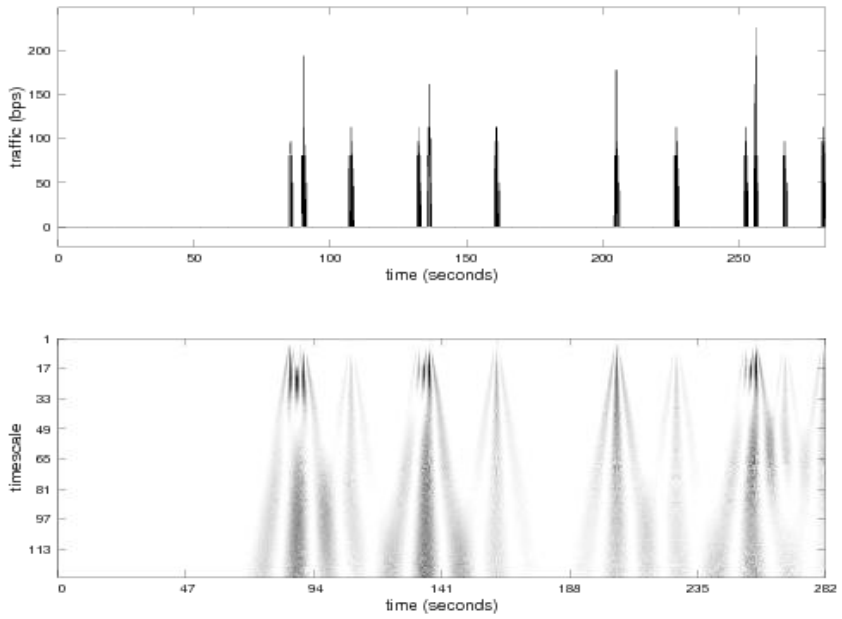

Figura 5.72 - Tráfego *downstream* IMAP por parte do servidor na direção A (bytes por segundo).

diminuta; logo nestes fluxos não há transmissão intensa de pacotes *downstream* em direção ao servidor. No segmento de altas frequências, os fluxos em que os clientes enviam mais informação para o servidor encontram-se nas regiões E e F.

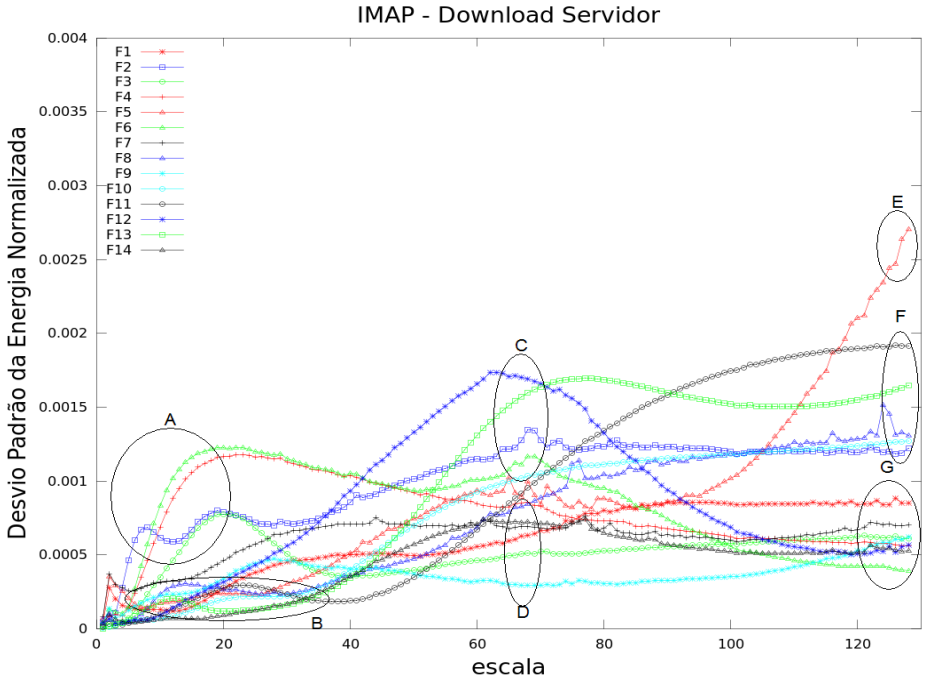

Figura 5.73 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* IMAP (do ponto de vista do servidor).

5.5 RTSP

5.5.1 Cliente (*Downstream*)

No que diz respeito ao tráfego *downstream* RTSP por parte do cliente, é possível observar as especificidades deste tráfego na Figura 5.74: transferência contínua de pacotes com picos de tráfego de amplitude semelhante, o que faz sentido se houver transmissão contínua de conteúdos, como por exemplo *streaming* de vídeo.

A Figura 5.75 apresenta algumas semelhanças com a Figura 5.74, apesar da transmissão de pacotes não ser contínua, o que se pode explicar por falhas na fonte de transmissão. De realçar que após estas falhas na transmissão, ocorrem picos de tráfego com amplitude superior aos restantes, que podem ser explicados pelo facto do utilizador efetuar um *refresh* do website para que a transmissão seja retomada. As componentes de média e alta frequência são mais intensas nestes picos de tráfego, o que dá consistência a esta alegação.

Analisando a Figura 5.76, é possível verificar a existência de vários fluxos de tráfego com eventos de muito baixa frequência com grande variação de energia, gerados por eventos pouco frequentes, como a sincronização automática e periódica da transmissão dos dados do servidor para o cliente e cliques do próprio cliente. A região B envolve eventos de baixa frequência com variação de energia considerável, normalmente associados a eventos em que o utilizador faz pedidos de novos conteúdos; no caso específico deste protocolo, pode corresponder ao caso em que o utilizador pretende escolher novas *streams* para visualizar.

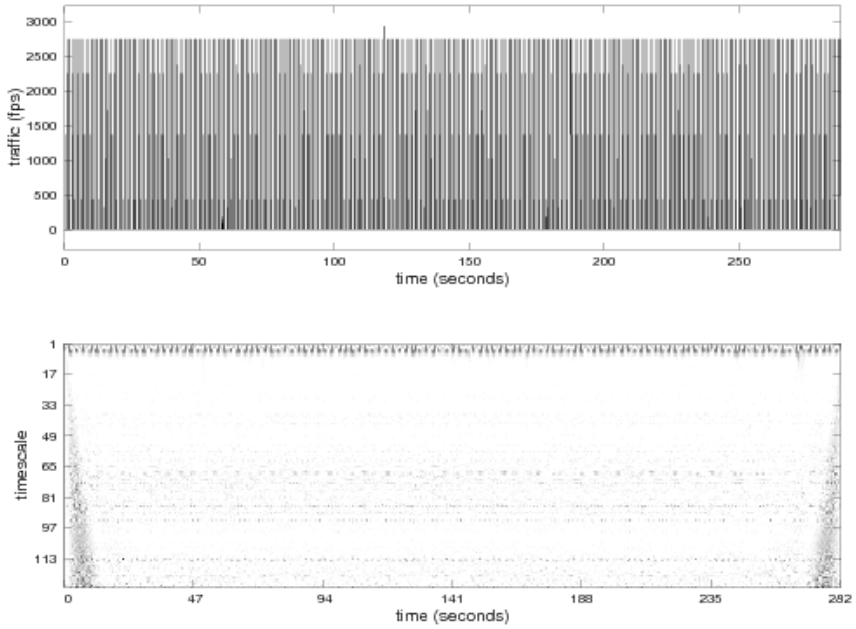

Figura 5.74 - Tráfego *downstream* RTSP por parte do cliente na direção B (bytes por segundo).

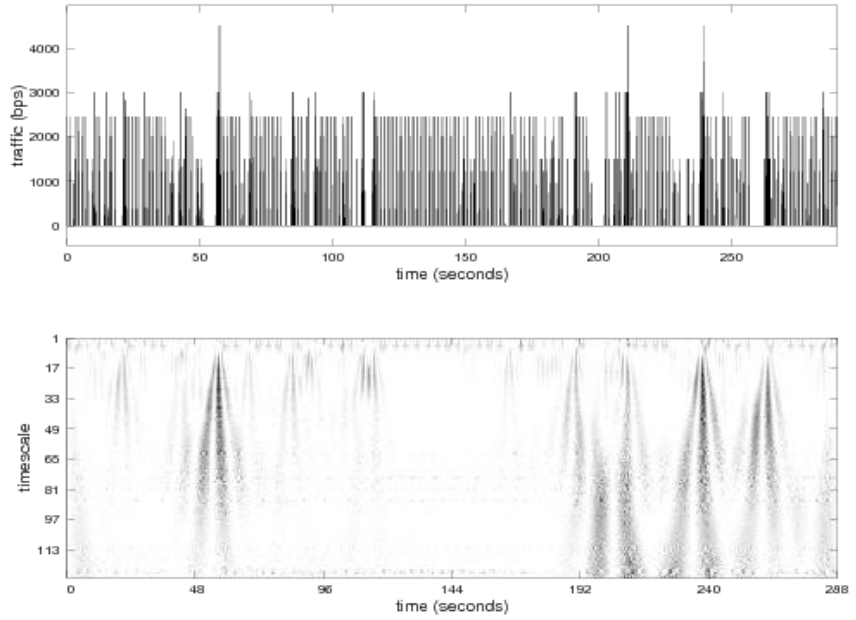

Figura 5.75 - Tráfego *downstream* RTSP por parte do cliente na direção A (bytes por segundo).

No segmento de médias frequências, os fluxos de tráfego estão divididos em duas regiões: na região C encontram-se os fluxos de tráfego com variação de energia considerável, o que indica mais eventos ligados a interações TCP e RTSP; variação de energia desta amplitude poderá estar ligada à abertura de sessões TCP e RTSP a pedido do cliente, para poder visualizar *streams*. A região F abrange eventos de alta frequência com percentagem reduzida de componentes de alta frequência; a presença de poucos eventos nesta região deve-se a uma transmissão reduzida de pacotes d*ownstream* para o cliente, o que não é muito comum se as transmissões *streaming* funcionarem em pleno. Pode então assumir-se neste caso que o conteúdo do *stream* foi transmitido durante

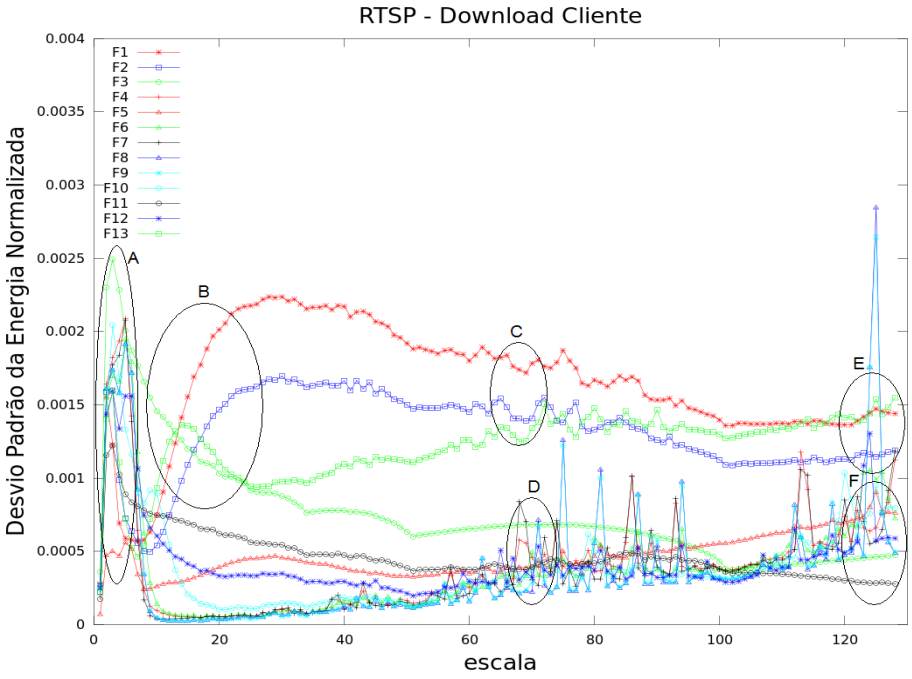

Figura 5.76 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* RTSP (do ponto de vista do cliente).

pouco tempo ou o cliente tentou aceder a determinados conteúdos para a sua visualização mas houve problemas com a ligação. No que diz respeito à região E, os eventos nela englobados apresentam uma variação de energia considerável, o que indica que a transmissão de pacotes *downstream* nestes fluxos é superior comparativamente à região E. É importante realçar o comportamento dos fluxos 8 e 9 que apresentam picos irregulares ao nível do desvio padrão da sua energia normalizada, principalmente em intervalos bem identificados: região A, segmento de médias frequências e segmento de frequências muito altas. Esta irregularidade pode ser explicada pelo facto da transmissão dos pacotes em situações de *streaming* de vídeo, por exemplo, não ser sempre constante, podendo existir períodos em que a transmissão de pacotes processa-se a grande velocidade – e o vídeo é imediatamente processado – como existem períodos em que a transmissão de pacotes é mais espaçada no tempo e assim a transmissão do vídeo pode sofrer cortes. As irregularidades na transmissão destes conteúdos ajudam a explicar estas variações instantâneas de energia em diferentes segmentos de frequência. [57]

5.5.2 Servidor (*Upstream*)

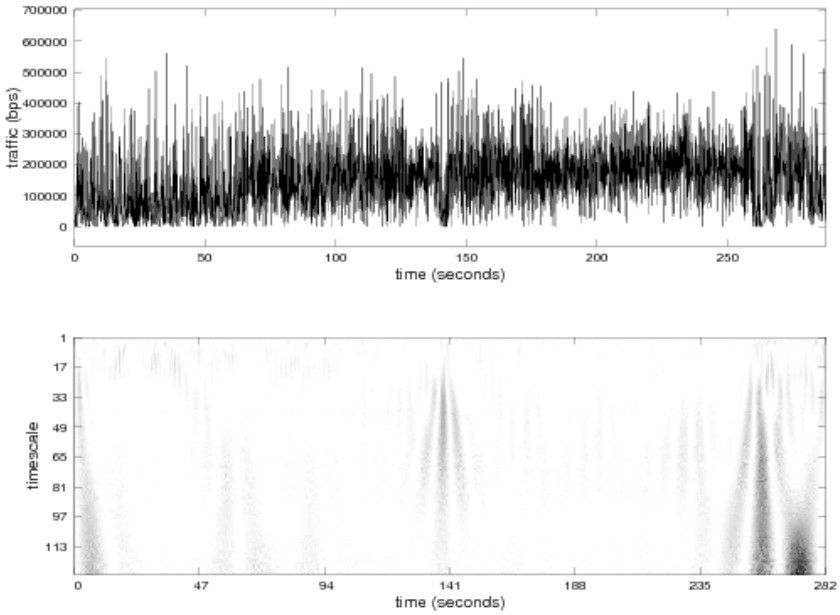

Figura 5.77 - Tráfego *upstream* RTSP por parte do servidor na direção B (bytes por segundo).

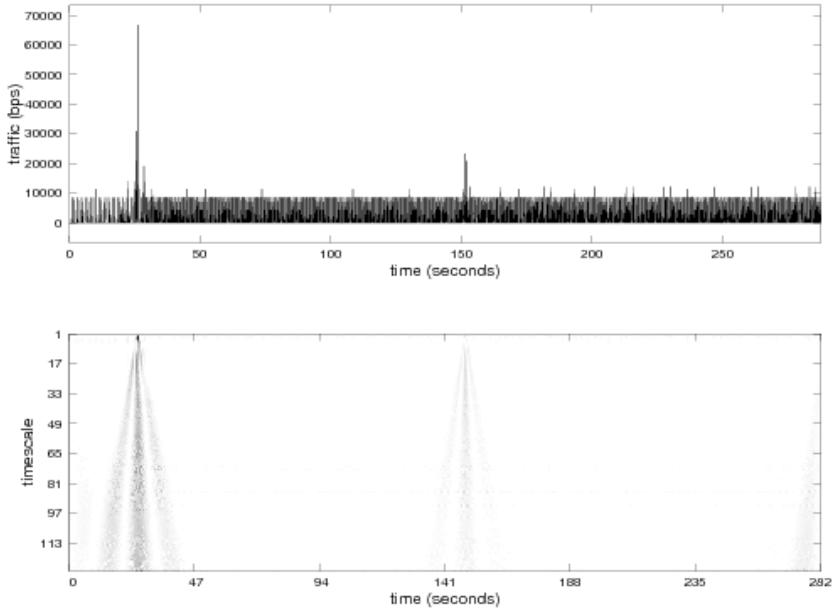

Figura 5.78 - Tráfego *upstream* RTSP por parte do servidor na direção B (bytes por segundo).

Para a situação de tráfego *upstream* RTSP por parte do servidor, foram analisados três casos distintos. A Figura 5.77 constitui exemplo da transmissão de conteúdos por *streaming*: tráfego contínuo de pacotes de grande tamanho e picos de tráfego de curta duração mas com amplitude muito elevada.

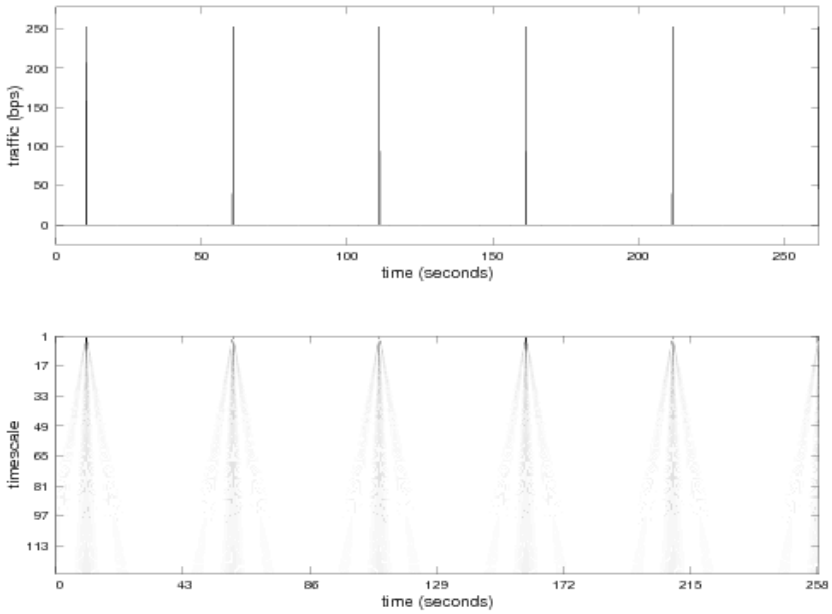

Figura 5.79 - Tráfego *upstream* RTSP por parte do servidor na direção A (bytes por segundo).

Analisando a Figura 5.78, verifica-se que após a ocorrência de um pico de tráfego no primeiro minuto o tráfego de pacotes intensifica-se, tornando-se homogéneo. O pico de tráfego inicial está associado a componentes de média e alta frequência, o que indicia a abertura de sessões para transmissão de conteúdos por *streaming.*

A Figura 5.79 constitui um exemplo onde não há transmissão por *streaming,* pois o tráfego é muito escasso (é enviado apenas um pacote de cada vez e muito separados temporalmente). O tráfego apresenta um carácter periódico, pois resulta de cliques do utilizador.

A análise da Figura 5.80 permite que as regiões demarcadas existentes nesta figura ocupem zonas semelhantes às regiões homónimas da Figura 5.76, com a ressalva que neste caso o tráfego é visto pelo ponto de vista do porto de partida, que neste caso pertence ao servidor. O fluxo 2 destaca-se por ter variações de energia consideráveis tanto na zona das baixas frequências como nas médias frequências, o que pressupõe que o cliente está a solicitar a abertura de novas páginas para poder escolher o conteúdo a visualizar, o que resulta em mais eventos e portanto maior variação de energia. Por outro lado, os fluxos 9 e 12 apresentam comportamento semelhante ao observado relativamente aos fluxos 8 e 9 da Figura 5.76, com variações instantâneas da energia em zonas bem definidas (região A, segmento de médias frequências e segmento de altas frequências), chegando a formar picos. A explicação para este fenómeno já foi dada na análise da Figura 5.76, sendo que agora as alterações ao nível da transmissão *upstream* dos pacotes para o cliente são observadas do ponto de vista do porto do servidor que envia os pacotes para o cliente.

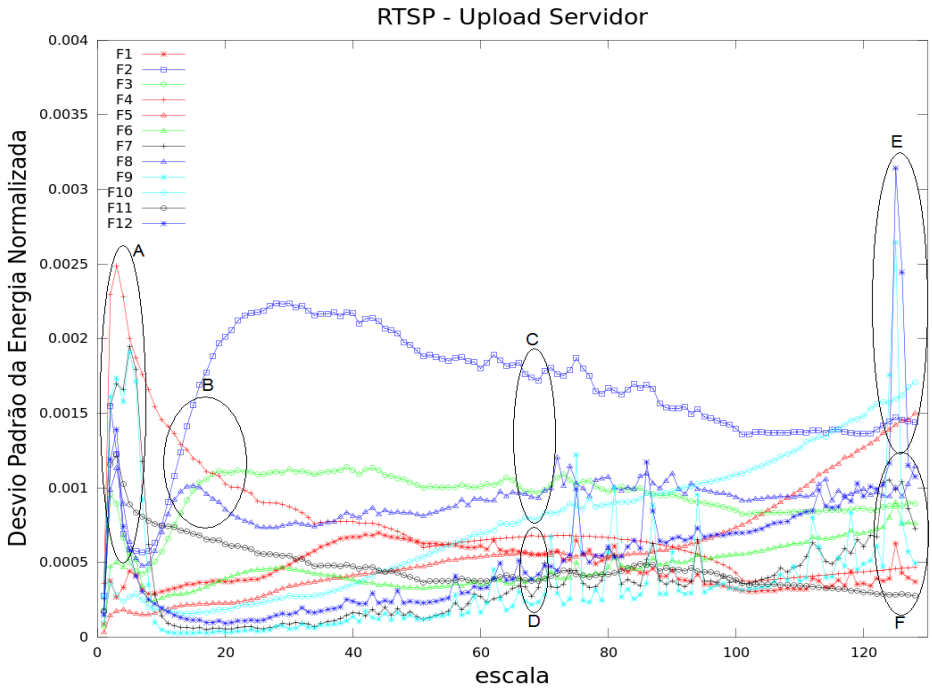

Figura 5.80 - Gráfico do desvio padrão da energia de vários fluxos de tráfego upstream RTSP (do ponto de vista do servidor).

5.5.3 Cliente (*Upstream*)

No que concerne ao tráfego *downstream* RTSP por parte do cliente, a Figura 5.81 apresenta tráfego contínuo de pacotes de pequeno tamanho. A inexistência de componentes de média e alta frequência no escalograma comprovam que este tráfego é de controlo da ligação entre o utilizador e o servidor. O facto de o tráfego ser contínuo indicia que estes pacotes são de resposta do cliente durante o *streaming* de conteúdos por parte do servidor.

A Figura 5.82 tem características semelhantes à figura anterior, à exceção da presença de picos de tráfego de baixa amplitude e curta duração e do tamanho dos restantes pacotes, que é menor comparativamente à Figura 5.81.

Relativamente à Figura 5.83, verifica-se que a região A engloba eventos de muito baixa frequência, gerados por ocorrências raras. Tendo em conta o cenário de transmissão de pacotes *upstream* por parte do cliente, é possível assumir que nos fluxos de tráfego considerados o cliente envia poucas solicitações para o servidor ou então os pacotes enviados são de tamanho diminuto. No segmento de médias frequências encontram-se duas regiões (B e C), sendo que a região B contém apenas um fluxo. Este fluxo consiste em eventos com variação de energia bastante superior aos eventos abrangidos pela região C, portanto neste fluxo de tráfego específico há mais interações TCP e RTSP e abertura de mais sessões TCP a pedido do cliente. Pode assumir-se que o cliente esteja a tentar aceder a várias *streams* (para selecionar a que tiver melhor qualidade, por exemplo) ou caso a ligação não seja a melhor, o cliente pode ser obrigado a atualizar a página do browser várias vezes. Relativamente ao segmento de altas frequências, observa-se que a grande maioria dos fluxos de tráfego analisados neste cenário encontram-se na região E e têm variação de energia pequena, o que faz sentido

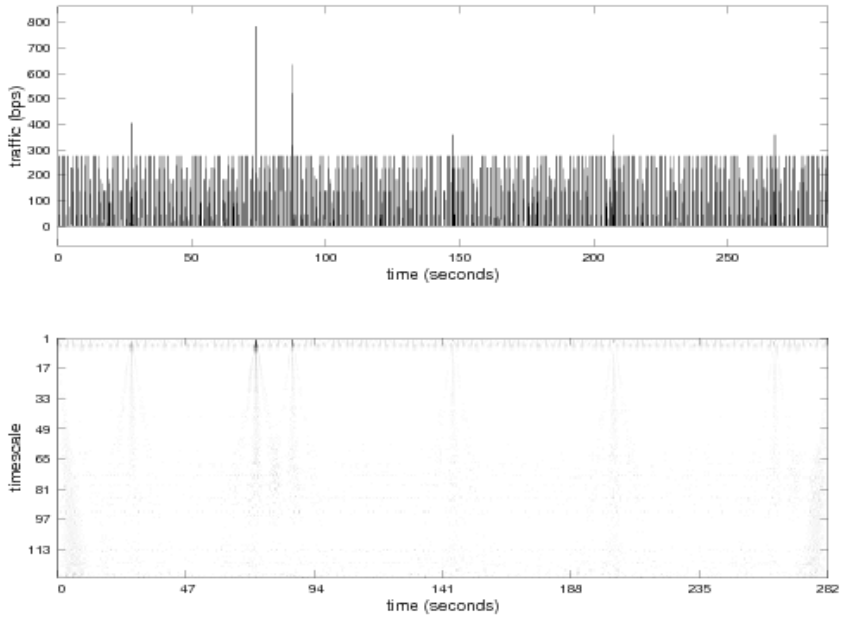

Figura 5.81 - Tráfego *upstream* RTSP por parte do cliente na direção B (bytes por segundo).

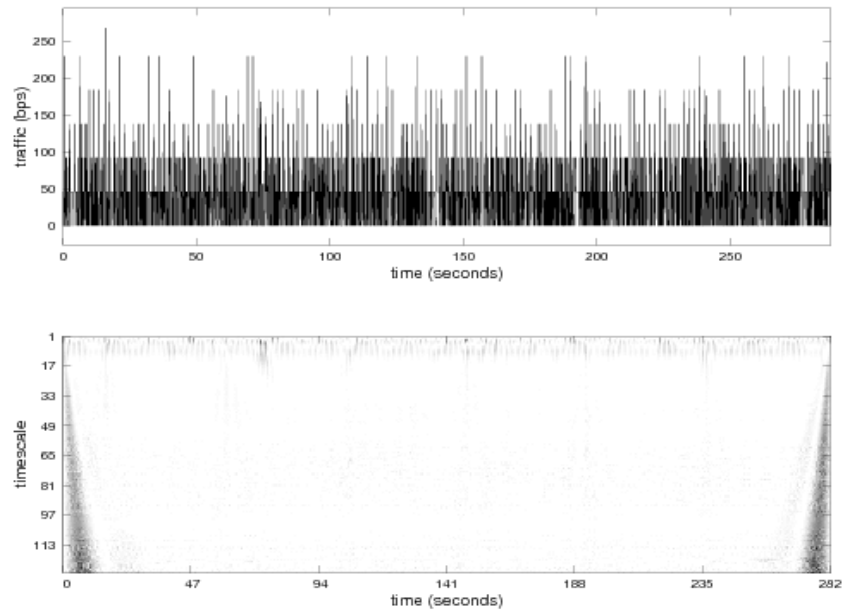

Figura 5.82 - Tráfego *upstream* RTSP por parte do cliente na direção B (bytes por segundo).

tendo em conta o cenário em questão. Os três fluxos presentes na região D contêm eventos com variação de energia superior aos eventos englobados na região E, o que significa uma maior transmissão de pacotes *upstream* por via do cliente. De realçar que os fluxos 3 e 9 apresentam tráfego com traçado irregular nos segmentos de médias e muito altas frequências, provavelmente devido à instabilidade da ligação entre cliente e servidor.

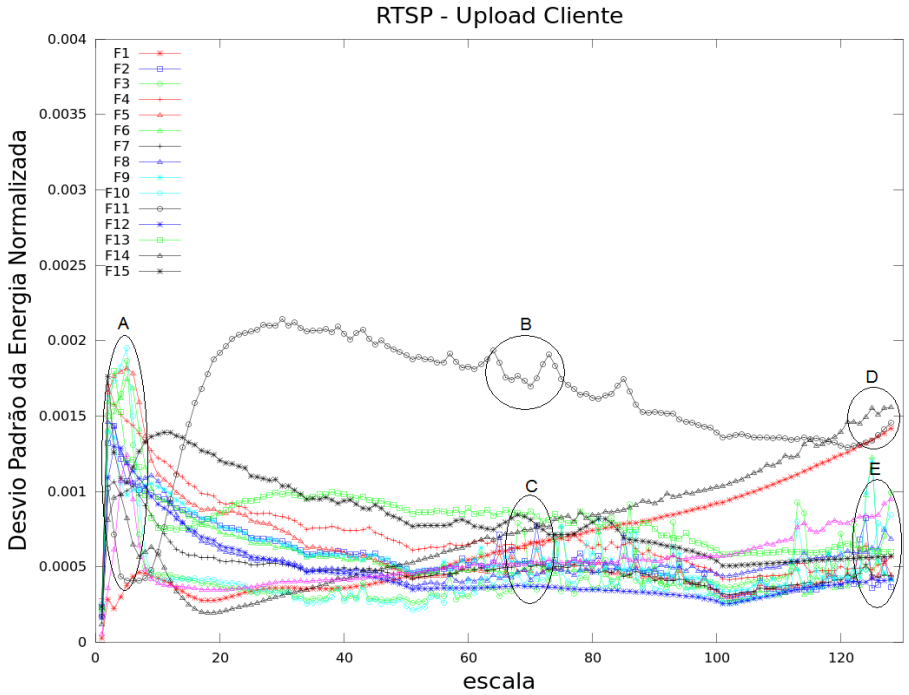

Figura 5.83 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* RTSP (do ponto de vista do servidor).

5.5.4 Servidor (*Downstream*)

A Figura 5.84 e a Figura 5.85 constituem exemplos dos pacotes de *acknowledge* enviados pelo utilizador para o servidor, durante a transmissão de uma *stream* (processo já abordado na secção 4.1.5). Os picos de tráfego presentes em ambos os casos representam comandos do utilizador com fim de alterar algum parâmetro da transmissão.

Relativamente à Figura 5.86, verifica-se que o tráfego é periódico e corresponde a cliques periódicos do utilizador, pois a frequência de chegada dos pacotes é pequena.

Analisando a Figura 5.87, verifica-se que o segmento de baixas frequências apresenta duas regiões distintas (apesar da região A apresentar eventos com as mesmas características da região homónima, anteriormente abordada na Figura 5.83). Quanto à região B, engloba eventos de baixa frequência com variação de energia reduzida, o que significa que nestes fluxos são enviados poucas solicitações por parte do cliente. A região C abrange todos os fluxos de tráfego considerados neste cenário e contêm eventos com pequena percentagem de componentes de média frequência, como esperado neste cenário. Tendo em conta que esta região envolve todos os fluxos de tráfego considerados neste cenário, é possível encontrar na mesma região fluxos associados a um número bastante variável de clientes.

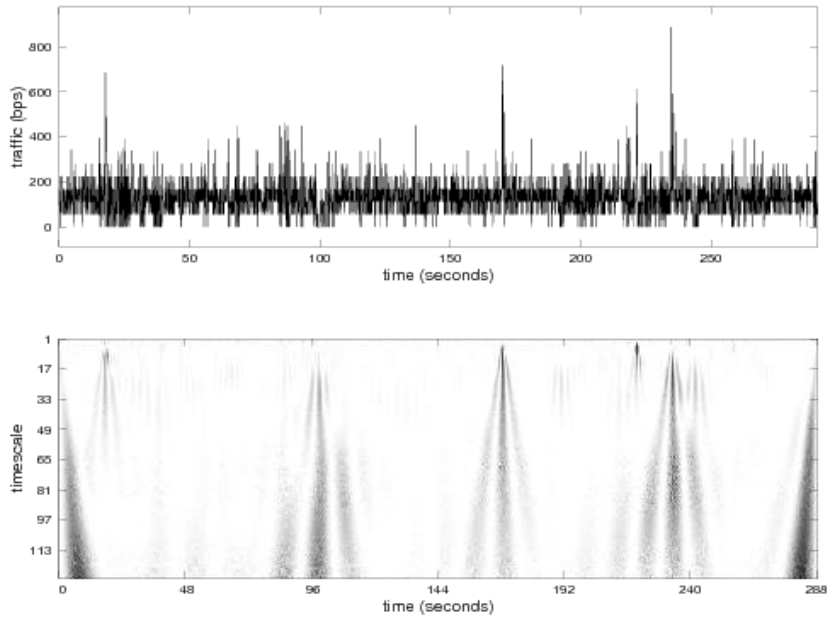

Figura 5.84 - Tráfego *downstream* RTSP por parte do servidor na direção A (bytes por segundo).

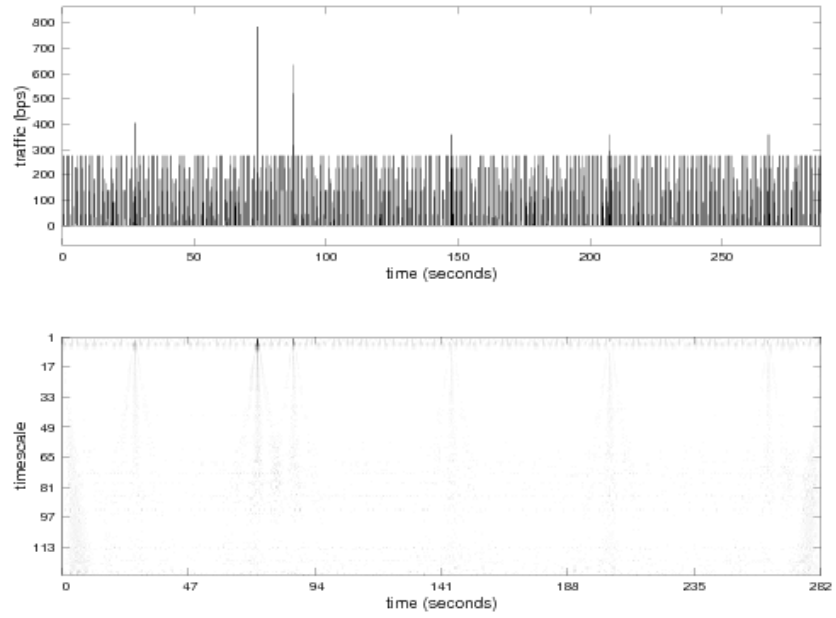

Figura 5.85 - Tráfego *downstream* RTSP por parte do servidor na direção B (bytes por segundo).

No segmento de altas frequências, apenas o fluxo 12 se destaca por apresentar uma percentagem considerável de componentes de alta frequência, o que indica tráfego intenso de pacotes *downstream* para o servidor e portanto atividade mais intensa por parte do(s) cliente(s) na gestão das *stream*s ou então o(s) cliente(s) está(ão) a ver conteúdos através de uma *stream* com uma taxa de transmissão grande e, como tal, gera mais pacotes de controlo e monitorização da ligação por parte do utilizador.

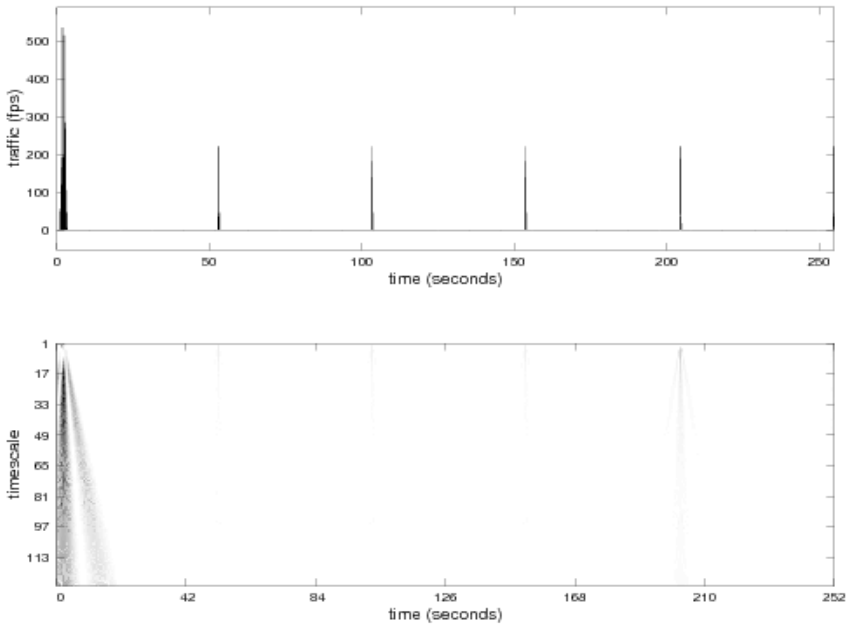

Figura 5.86 - Tráfego *downstream* RTSP por parte do servidor na direção B (bytes por segundo).

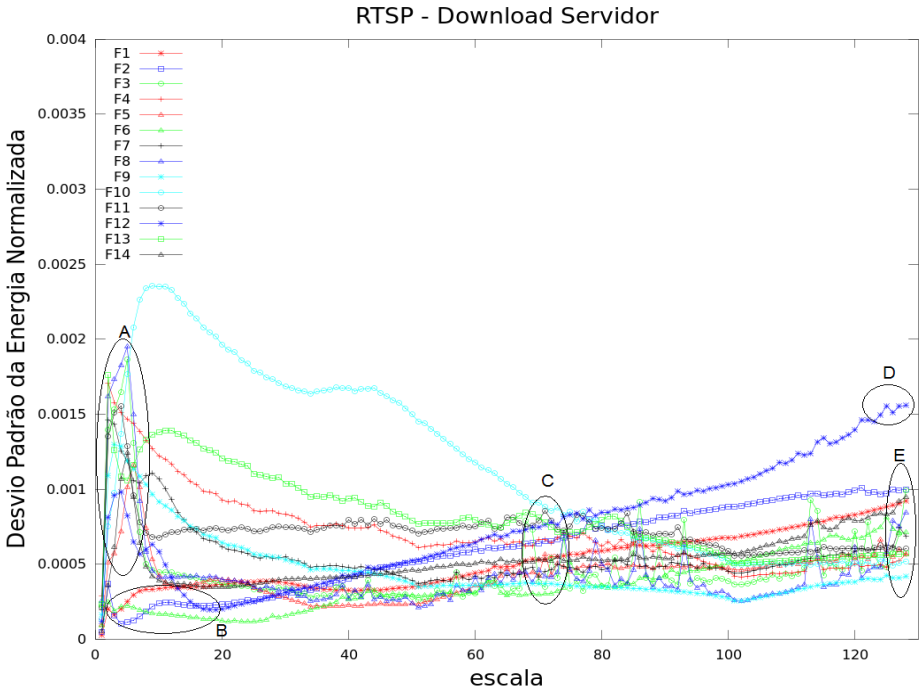

Figura 5.87 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* RTSP (do ponto de vista do servidor).

5.6 MSNP

5.6.1 Cliente (*Downstream*)

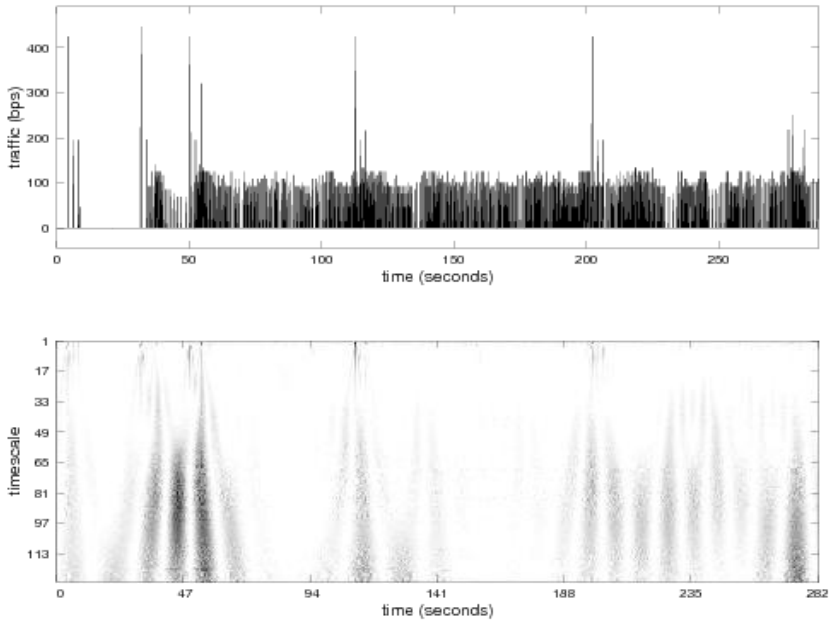

Figura 5.88 - Tráfego *downstream* MSNP por parte do cliente na direção A (bytes por segundo).

A análise do tráfego *downstream* MSNP por parte do cliente é relativamente acessível, pois a grande diferença entre os dois casos em análise reporta-se à frequência da chegada de pacotes. Na Figura 5.88, há chegada de pacotes de pequeno tamanho, registando-se a ocorrência de picos de tráfego ocasionalmente. Estes picos de tráfego têm curta duração e pequena amplitude e geram componentes de alta frequência de pouca intensidade, pois geralmente o tráfego recebido nas sessões MSNP consiste sobretudo em mensagens de texto. O restante tráfego trata-se de mensagens de controlo da aplicação (aviso de escrita, notificações de presença, entre outras).

Já na Figura 5.89 o tráfego não é contínuo como na figura anterior, mas continuam a verificar-se os picos de tráfego (correspondentes às mensagens de texto enviadas), que precedem sempre a chegada de novos pacotes. Os pacotes mais pequenos correspondem às mensagens de controlo e notificações da aplicação. Este caso pode ser o exemplo de uma conversa de chat entre dois utilizadores com vários momentos de pausa e daí os intervalos de tempo sem qualquer tráfego. O tráfego em ambas as figuras é recebido na porta de serviço MSNP, o que significa que o tráfego analisado nestas duas figuras poderá incluir pacotes de controlo da ligação (principalmente se os clientes nos terminais da ligação estiverem conectados a servidores diferentes) para além das mensagens que os clientes trocam entre si.

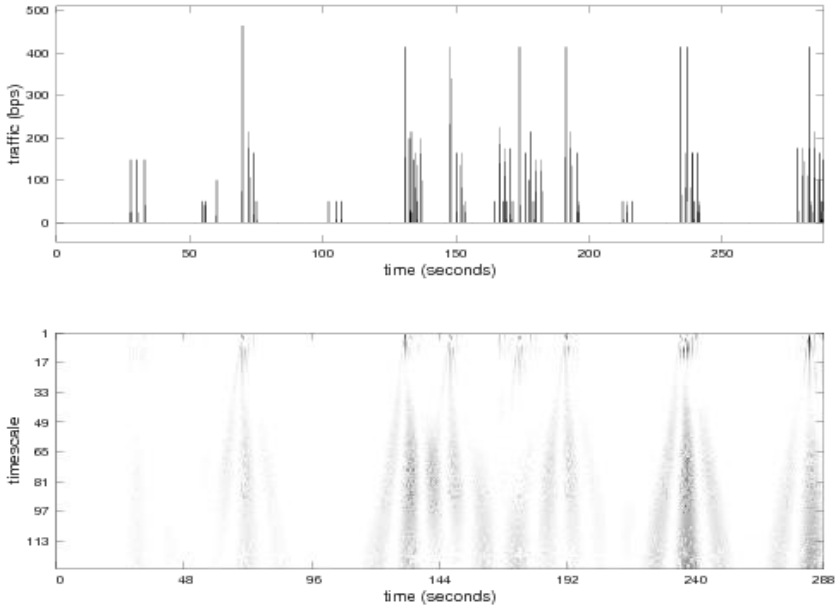

Figura 5.89 - Tráfego *downstream* MSNP por parte do cliente na direção A (bytes por segundo).

Analisando a Figura 5.90, verifica-se que no segmento de baixas frequências existe apenas a região A que engloba todos os fluxos de tráfego envolvidos neste cenário. Os eventos localizados nesta região apresentam uma frequência muito baixa, pois são gerados por acontecimentos com carácter periódico, o que faz sentido nesta aplicação se os intervalos de tempos entre a escrita e a resposta tiverem uma duração semelhante. No segmento de médias frequências, verifica-se que todos os fluxos (à exceção do fluxo 4) encontram-se na região C e apresentam variação de energia baixa, pois o envio de mensagens de texto para outros utilizadores online cria poucas sessões e as interações UDP e MSNP são de pouca monta. O fluxo 4 destaca-se, pois os seus eventos mostram uma variação de energia considerável e superior comparativamente ao observado na região B. Esta ocorrência pode dever-se ao facto do cliente estar a conversar com outros utilizadores da mesma aplicação ao mesmo tempo, resultando daí uma maior interação de tráfego MSNP e UDP. No segmento de altas frequências, observa-se que grande parte dos fluxos analisados neste cenário encontram-se na região E e apresentam uma taxa de chegada de pacotes reduzida, o que era esperado tendo em conta que as mensagens de texto normalmente enviadas nesta aplicação requerem pacotes d pequeno tamanho para serem entregues no seu destino. Os dois fluxos que fogem à regra (fluxos 4 e 7) situam-se na região D, em que a variação de energia dos eventos gerados é ligeiramente superior aos eventos da região E, pois a taxa de chegada de pacotes é maior. Isto pode ser explicado se o cliente encetar conversas em simultâneo com outros utilizadores da mesma aplicação.

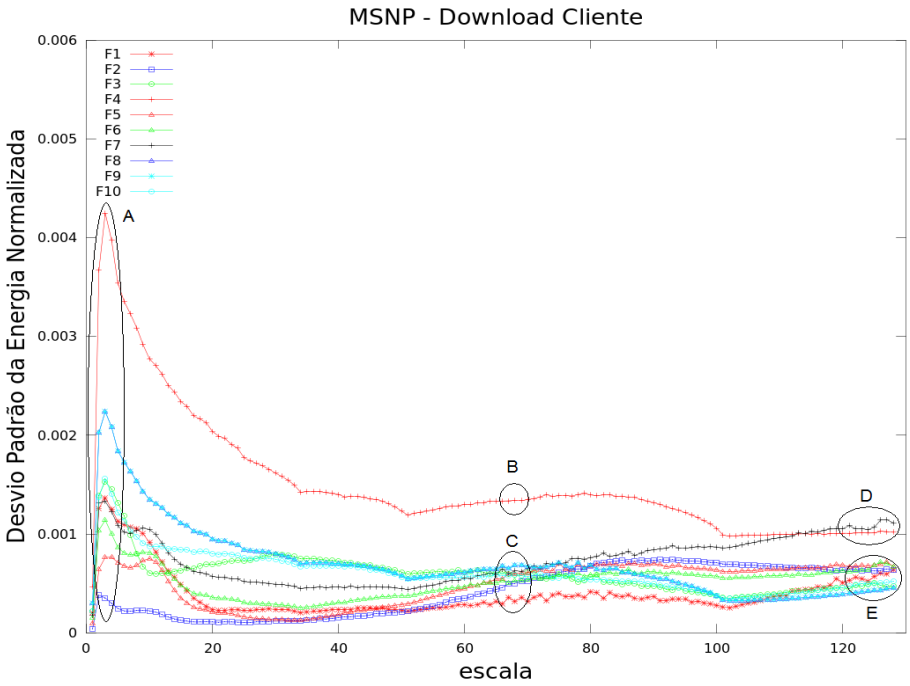

Figura 5.90 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* MSNP (do ponto de vista do cliente).

5.6.2 Servidor (*Upstream*)

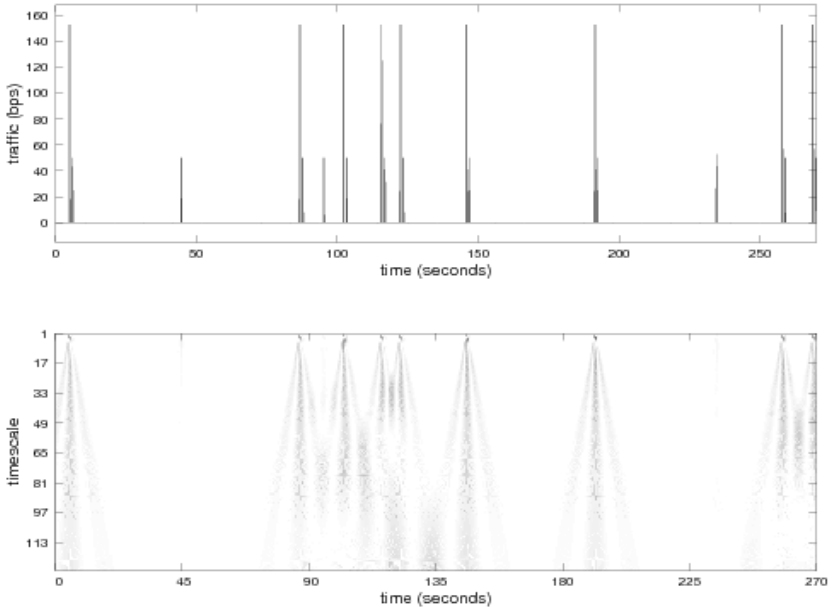

Figura 5.91 - Tráfego *upstream* MSNP por parte do servidor na direção A (bytes por segundo).

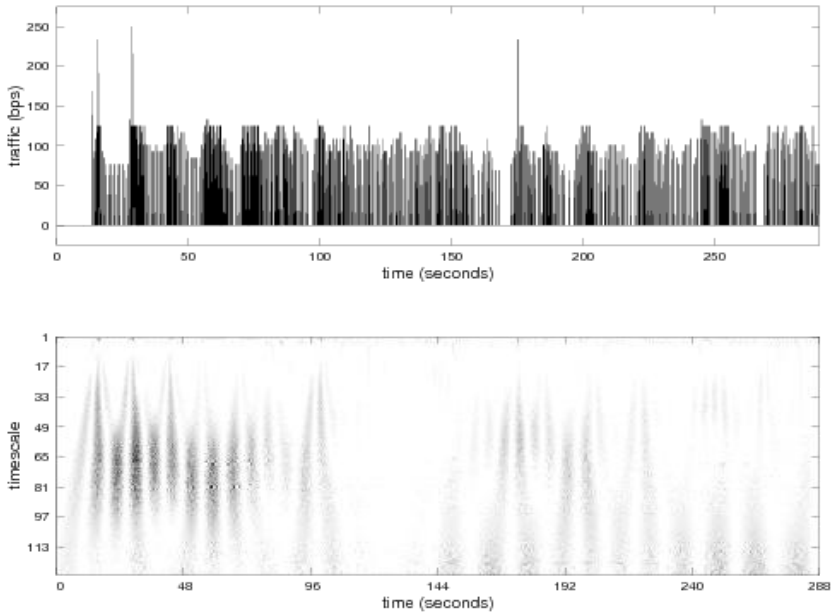

Figura 5.92 - Tráfego *upstream* MSNP por parte do servidor na direção A (bytes por segundo).

A Figura 5.91 e a Figura 5.92 constituem dois exemplos possíveis para cenários de chat online: no primeiro exemplo, tem-se tráfego de pacotes de pequena dimensão separados por intervalos de tempo que podem chegar às dezenas de segundos associado a componentes de baixa e média frequência no escalograma; no segundo exemplo o tráfego é muito mais intensivo, ocorrendo ocasionalmente picos de tráfego e praticamente não tem paragens; está associado a componentes de média frequência, Pode dizer-se que nesta segunda figura estão presentes muitos clientes, pois o tráfego é muito intenso e tem aspecto praticamente constante, que resulta da soma de diferentes fluxos.

A análise da Figura 5.93 permite verificar que à exceção das regiões C e D, a descrição das restantes regiões é similar à descrição das regiões homónimas efetuada na Figura 5.90, com a ressalva que neste caso o tráfego é analisado do ponto de vista do servidor. O fluxo 10 demarca-se dos restantes pelo seu comportamento nos segmentos de médias e altas frequências, apresentando eventos com grande variação de energia em ambos os segmentos. Este comportamento pode ser explicado se o cliente interagir com vários utilizadores simultaneamente, trocando mensagens de texto e porventura efetuando vídeo chamadas e troca de ficheiros, o que originaria uma maior taxa de transmissão de pacotes a partir do servidor para poder servir esta procura.

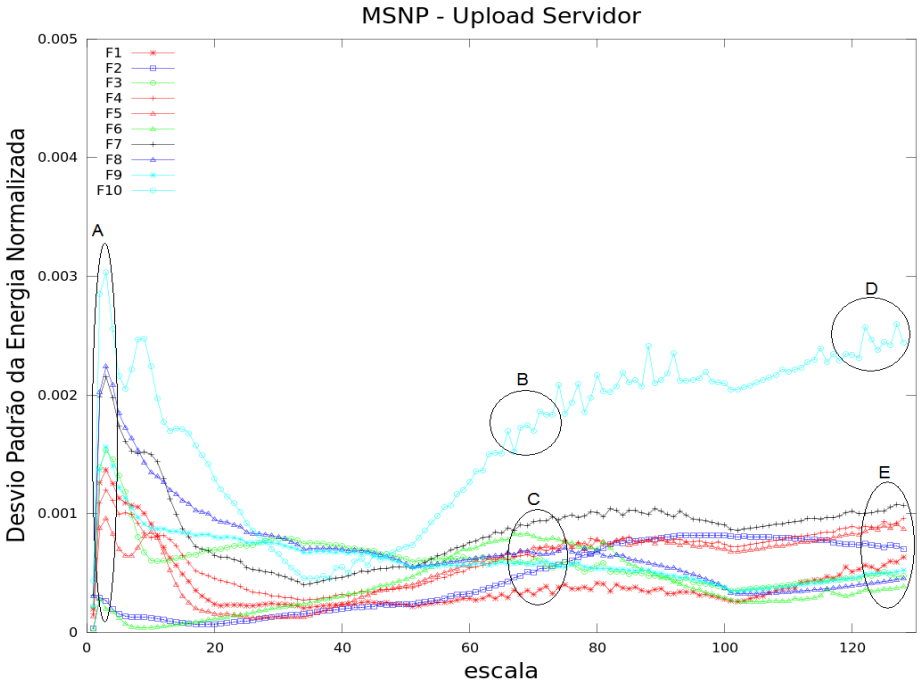

Figura 5.93 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* MSNP (do ponto de vista do servidor).

5.6.3 Cliente (*Upstream*)

Na Figura 5.94, observa-se que existe tráfego contínuo de pacotes durante grande parte do intervalo de tempo considerado, com um pico de tráfego no início da dita sequência. Tendo em conta que o tráfego MSNP trata-se principalmente do envio de mensagens de texto, é possível observar no escalograma que as componentes de média e alta frequência apresentam pouca intensidade, exceto no pico de tráfego inicial. Esse pico de tráfego gera componentes de frequência relevantes pois é responsável pelo início da sessão. Os picos de tráfego de média amplitude são gerados pelo envio de mensagens de texto e o restante tráfego tem um perfil constante (são gerados por mensagens de controlo ou outras notificações).

Na Figura5.95, o tráfego apresenta-se não periódico e com picos de tráfego de moderada duração e com amplitude baixa, mas semelhante a todos os pacotes. Tendo em conta que estes picos de tráfego geram componentes de média frequência com alguma intensidade, pode assumir-se que o utilizador está a criar diferentes sessões UDP, interligando-se com diferentes utilizadores online. Nestas duas figuras em questão, o tráfego capturado é enviado pela porta de serviço MSNP, portanto volta a colocar-se a questão já abordada no cenário de tráfego *downstream* para o cliente (secção 5.6.1.).

Analisando a Figura 5.96, verifica-se que os fluxos localizados na região A apresentam o mesmo comportamento dos fluxos presentes na região homónima na Figura 5.93, embora neste caso o tráfego de pacotes *upstream* seja proveniente do cliente e nesse cenário o tráfego de pacotes *upstream* provenha do servidor. O segmento de médias frequências encontra-se dividido em duas regiões, sendo que a região C engloba fluxos de tráfego com baixa variação de energia e a região B abrange fluxos com variação de energia moderada; nesta última região, a variação de energia dos eventos em causa será maior por haver mais interações MSNP e UDP, ou seja, nestes fluxos o cliente interage com mais utilizadores da mesma aplicação comparativamente

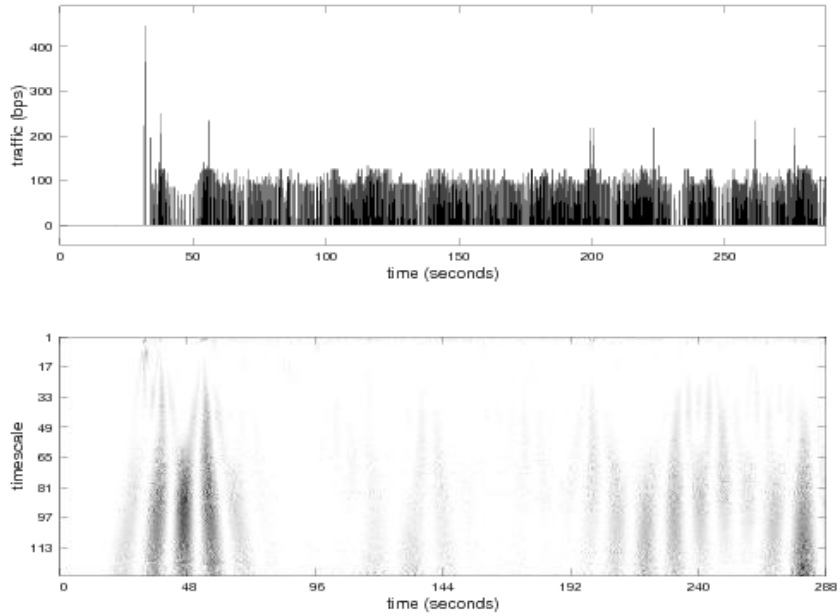

Figura 5.94 - Tráfego *upstream* MSNP por parte do cliente na direção A (bytes por segundo).

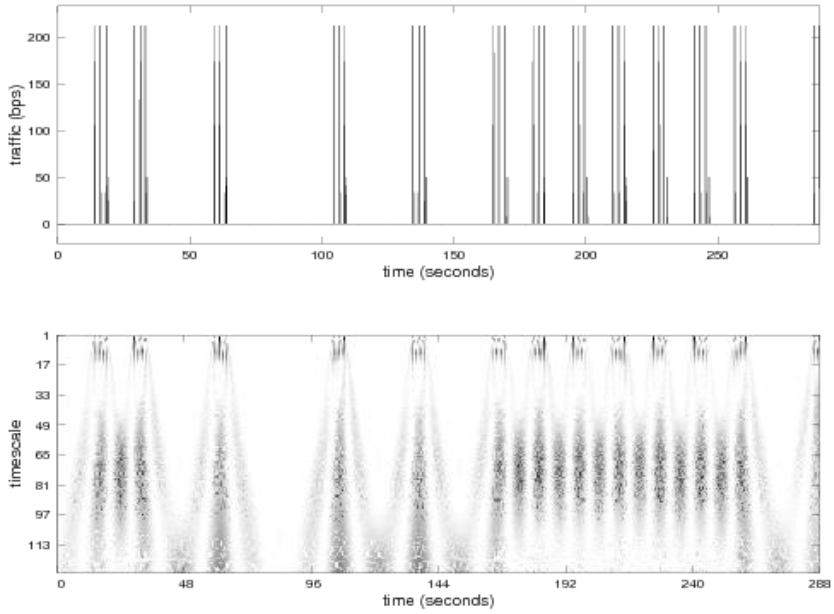

Figura5.95 - Tráfego *upstream* MSNP por parte do cliente na direção A (bytes por segundo).

aos fluxos da região C. No segmento de altas frequências observa-se que, à exceção do fluxo 10, os restantes fluxos encontram-se na região E. Nesta região as taxas de transmissão de pacotes são normalmente pequenas, o que sugere que o cliente interage com poucos utilizadores da mesma aplicação. No que respeita à região D, esta apenas inclui o fluxo 10; os eventos associados a este fluxo são responsáveis por uma taxa de transmissão de pacotes bastante considerável por parte do cliente, o que significa

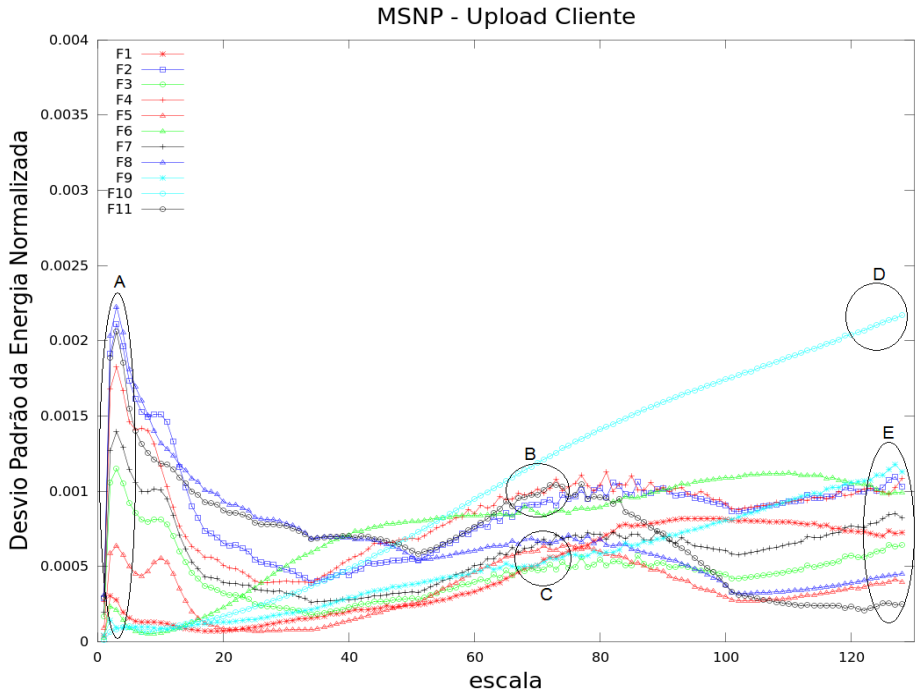

Figura 5.96 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* MSNP (do ponto de vista do cliente).

que o cliente em causa interage com de forma bastante ativa, daí o tráfego intenso que ele gera.

5.6.4 Servidor (*Downstream*)

Analisando a Figura 5.97, é possível observar que o tráfego tem um carácter não periódico e que os componentes de frequência gerados por cada pico de tráfego têm um perfil similar, pois todos eles geram componentes de baixa e média frequência, o que é expectável tendo em conta que o tráfego relacionado com *chatting* consiste maioritariamente em mensagens de texto.

A Figura 5.98 apresenta tráfego contínuo durante grande parte do intervalo temporal considerado, incluindo alguns picos de tráfego de curta duração e baixa amplitude. Estes picos de tráfego são gerados por cliques do utilizador, o que é validado pelos componentes de frequência que surgem no escalograma associados a estes mesmos picos de tráfego; são sobretudo componentes de média e alta frequência de média amplitude, o que tendo em conta o tipo de tráfego envolvido é aceitável. Também é possível vislumbrar pequenos componentes de baixas frequências, diretamente associados à atividade do utilizador. Exceptuando os picos de tráfego antes referidos, o restante tráfego tem um perfil constante, correspondente a mensagens de controlo e a notificações da aplicação. Nestas duas figuras em questão, o tráfego capturado é recebido pela porta de serviço MSNP, portanto volta a colocar-se a questão já abordada no cenário de tráfego *downstream* para o cliente (secção 5.6.1.) e tráfego *upstream* oriundo do cliente.

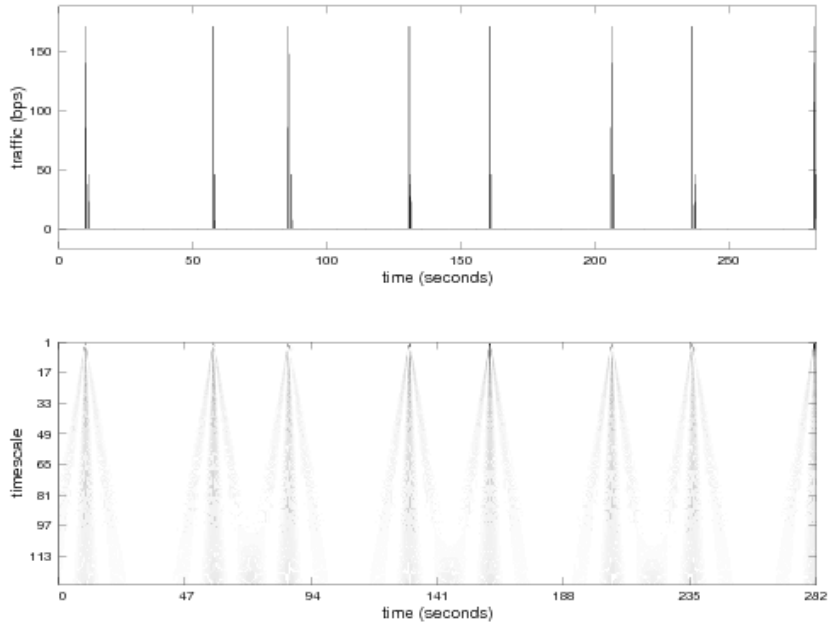

Figura 5.97 - Tráfego *downstream* MSNP por parte do servidor na direção B (bytes por segundo).

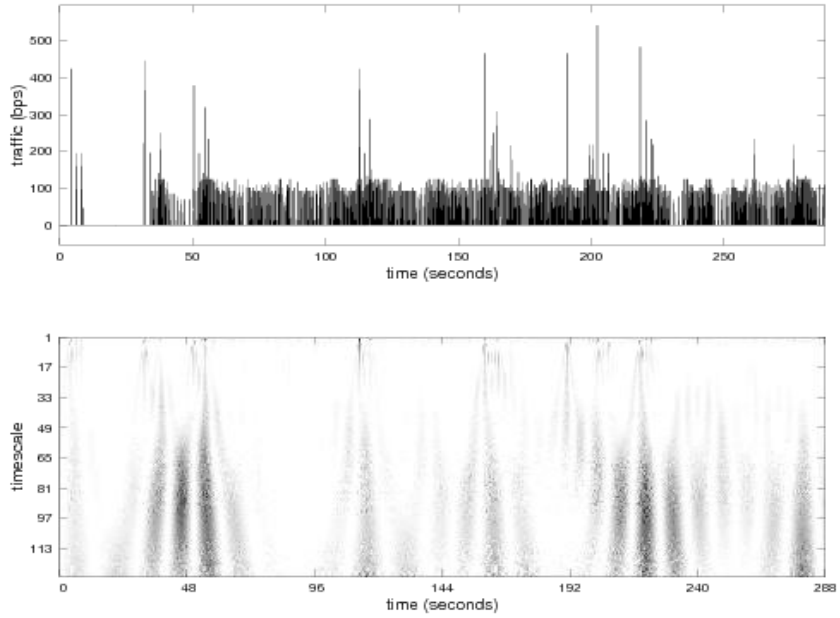

Figura 5.98 - Tráfego *downstream* MSNP por parte do servidor na direção A (bytes por segundo).

Analisando a Figura 5.99, observa-se que todos os fluxos de tráfego estão abrangidos na mesma região em cada segmento de frequência analisado, e que portanto os fluxos de tráfego *downstream* MSNP têm um comportamento bastante similar. Assim, a região A abrange eventos de muito baixa frequência, gerados por cliques periódicos do cliente, o que se explica se o tempo de escrita e resposta não sofrer grandes modificações ao longo do tempo. No segmento de médias frequências verifica-se que

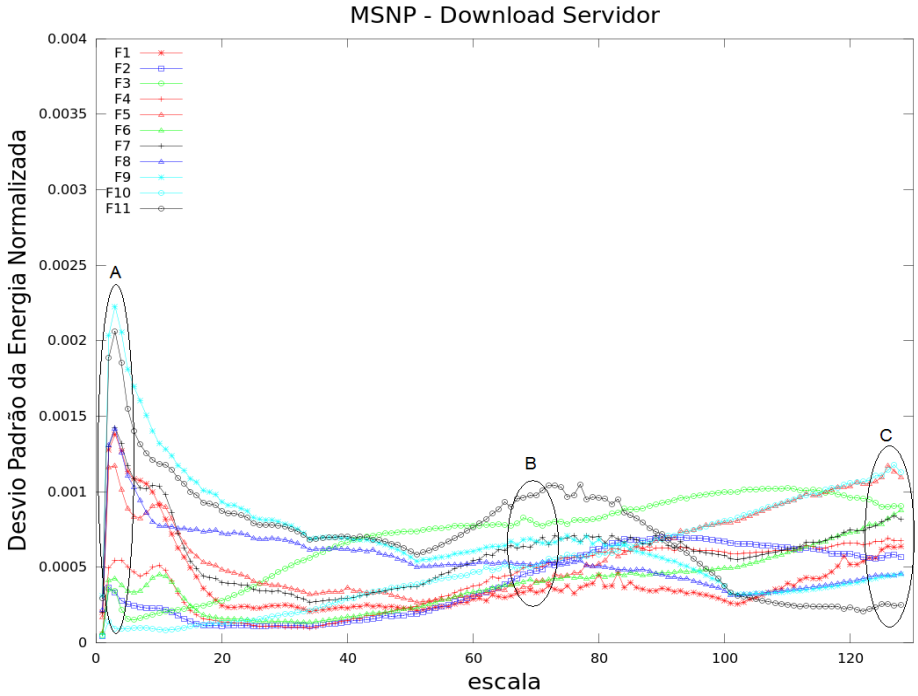

Figura 5.99 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* MSNP (do ponto de vista do servidor).

estes fluxos são responsáveis por poucas interações MSNP e UDP como era esperado. Finalmente no segmento de altas frequências, os fluxos encontram-se numa faixa de variação de energia com pouca amplitude, o que indicia que a taxa de transmissão de pacotes por parte do cliente é baixa e, portanto, este não está a encetar conversas com muitos contactos ao longo do tempo e as suas conversas não envolvem troca de ficheiros anexos nem vídeo chamadas.

5.7 XBOX

5.7.1 Cliente (*Downstream*)

A análise do download de tráfego XBOX requer alguns cuidados, pois como foi explicado na secção 4.1.7,o serviço *XBOX Live* disponibiliza conteúdos como filmes e séries, assim como o suporte para jogos online. Contudo, o porto 3074 está atribuído aos jogos da XBOX, portanto estes conteúdos serão à partida transmitidos através do porto 80 (HTTP). A atividade de um utilizador é muito mais intensa durante um jogo comparativamente ao *browsing*. Assim, a resposta por parte do servidor aos comandos do utilizador terá de ser rápida, requerendo portanto grande largura de banda, proporcional à quantidade de jogadores online em simultâneo. Outro aspeto importante prende-se com os diferentes tipos de jogos disponíveis neste serviço: as características da atividade do utilizador durante um jogo de estratégia são distintas das características apresentadas durante um jogo de desporto, por exemplo.

A Figura 5.100 apresenta características de tráfego semelhantes às normalmente associadas à visualização de vídeos online [44]. Contudo, o tráfego tem um perfil muito

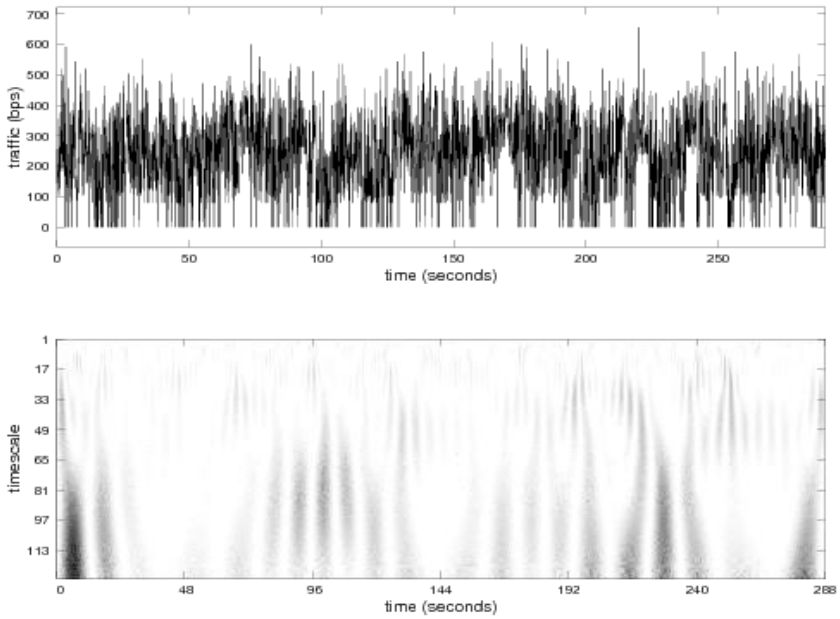

Figura 5.100 - Tráfego *downstream* XBOX por parte do cliente na direção A (bytes por segundo).

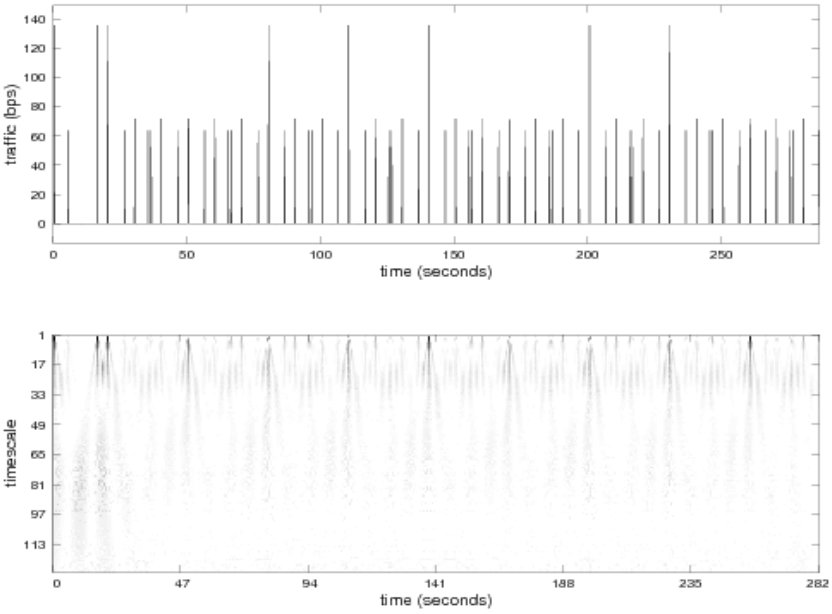

Figura 5.101 - Tráfego *downstream* XBOX por parte do cliente na direção A (bytes por segundo).

irregular, pois durante os jogos o ambiente de jogo é influenciada pelos comandos do jogador, o que ajuda a explicar a existência de componentes de alta frequência com amplitude considerável.

Em relação à Figura 5.101, realça-se o facto de haver picos de tráfego bem definidos, associados a componentes de baixa frequência gerados pelos cliques do utilizador. Tendo em conta que no intervalo temporal considerado o tráfego não é

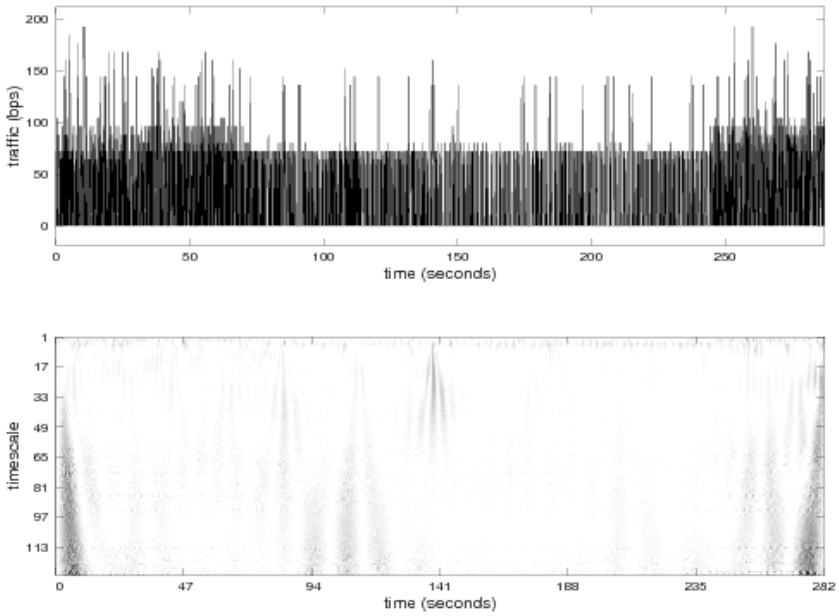

Figura 5.102 - Tráfego *downstream* XBOX por parte do cliente na direção B (bytes por segundo).

contínuo e os picos de tráfego têm períodos de tempo relativamente extensos a separá-los, pode assumir-se que o utilizador esteja a jogar um jogo que não requer atividade constante, como são exemplo os jogos de estratégia.

Relativamente à Figura 5.102, verifica-se que o tráfego é contínuo, com múltiplos picos de tráfego de curta duração e amplitudes relativamente pequenas. Estes são responsáveis pelo surgimento de pequenos componentes de baixa frequência no escalograma, mas tendo em conta a proximidade temporal entre os diferentes picos, é possível assumir que esta figura representa um caso de atividade mais intensa do utilizador comparativamente à Figura 5.101.

Analisando a Figura 5.103, encontram-se duas regiões no segmento de baixas frequências. A região A engloba eventos de muito baixa frequência, que são gerados por acontecimentos que se registam raramente. Neste caso, estes acontecimentos podem ser a atividade do utilizador na interface do serviço *XBOX Live* assim como sincronizações automáticas entre o servidor remoto do serviço e o terminal do cliente para assegurar a qualidade da ligação. A região B envolve eventos de baixa frequência com variação de energia de pequena amplitude. A região C situa-se no segmento de médias frequências e abrange todos os fluxos de tráfego, pois a variação de energia não difere muito de fluxo para fluxo nesta região. Assim, a variação de energia destes fluxos é normalmente de baixa amplitude, o que indicia a criação de poucas sessões UDP durante o intervalo de tempo em que o cliente se encontra a jogar. No segmento de altas frequências encontram-se três regiões bem definidas tendo em conta a taxa de transmissão de pacotes: na região F, a taxa de receção de pacotes por parte do cliente é muito baixa; na região E, a taxa de receção de pacotes por parte do cliente é considerável, pelo menos comparativamente à região F; finalmente, na região D a taxa de receção de pacotes por parte do cliente é bastante grande (fluxos 4,8,9 e 18). É possível assumir que os fluxos da região D estejam associados a jogos em que o cliente tenha de ser mais ativo como os jogos de ação na primeira pessoa (*First Person Shooter*), jogos de desporto ou de aventura. Já os fluxos da região E, com taxa receção de pacotes mais comedida, poderão estar a jogos de pergunta/resposta e até certos jogos de estratégia.

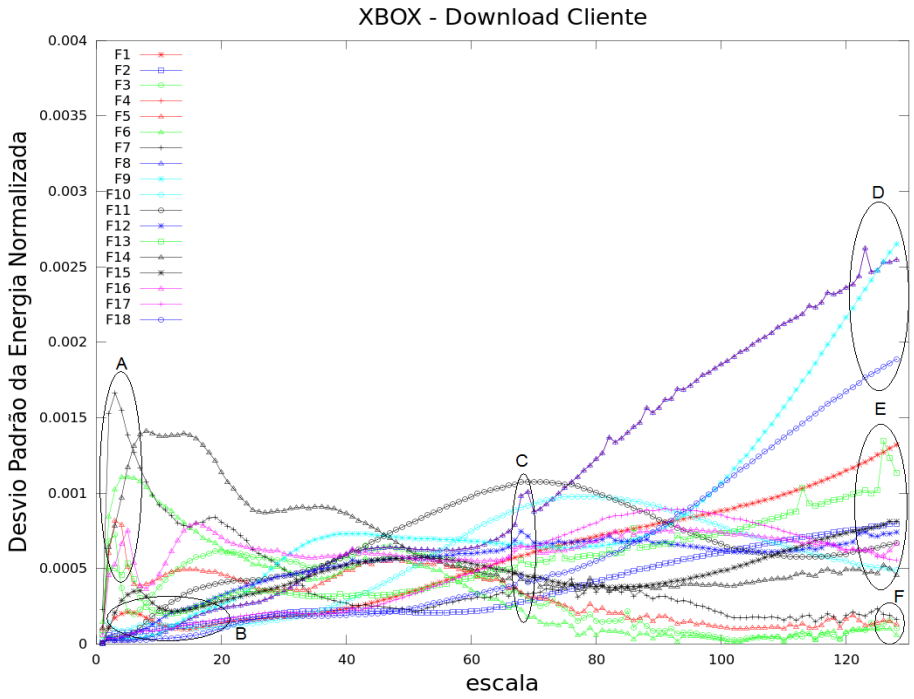

Figura 5.103 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* XBOX (do ponto de vista do cliente).

5.7.2 Servidor (*Upstream*)

Analisando o tráfego visto do servidor da Figura 5.105, verifica-se que apresenta um perfil muito idêntico ao da Figura 5.100, com a devida ressalva que esta figura reportava-se ao tráfego *downstream* por parte do cliente, ou seja, o tráfego é analisado a partir do porto de destino enquanto na Figura 5.105, o tráfego é visto a partir do porto de origem do mesmo.

A Figura 5.104 tem características muito semelhantes às da Figura 5.105, à exceção dos últimos quarenta segundos em que ocorre uma redução abrupta do tráfego. Até esse momento o tráfego mantinha-se contínuo e com carácter não periódico, mas após um último pico de tráfego esta cadência termina e o tráfego passa a ser periódico mas com pacotes de pequeno tamanho a circularem com espaço de alguns segundos entre eles. Tendo em conta os dados em análise, é possível assumir-se que o utilizador estaria a jogar um jogo e interrompeu a sua atividade.

Na Figura 5.106 é possível identificar um pico de tráfego de curta duração e baixa amplitude gerado por cliques do utilizador (observando o escalograma, está associado a componentes de baixa frequência). O restante tráfego tem um carácter periódico e é constituído por pacotes de pequenas dimensões.

Analisando a Figura 5.107, observa-se que as regiões do segmento de baixas frequências (regiões A e B) têm características similares às suas regiões homónimas do cenário de fluxos de tráfego *downstream* XBOX em direção do cliente, sendo que neste caso os fluxos são observados do ponto de vista do servidor, que é a fonte desse mesmo tráfego. No segmento de médias frequências encontram-se duas regiões, que se distinguem pela variação de energia dos fluxos envolvidos em cada uma: a região C engloba eventos com variação de energia de amplitude considerável, o que implica a

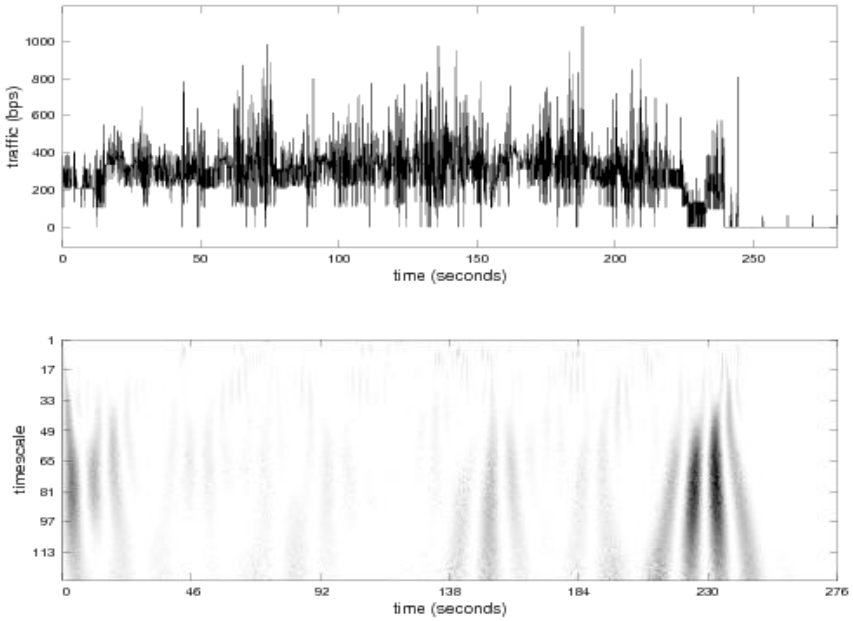

Figura 5.104 - Tráfego *upstream* XBOX por parte do servidor na direção B (bytes por segundo).

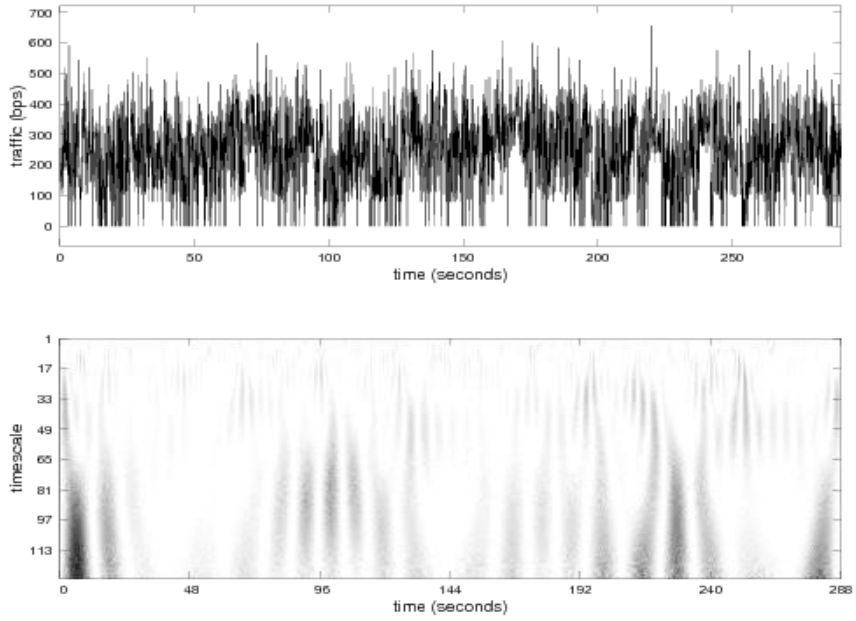

Figura 5.105 - Tráfego *upstream* XBOX por parte do servidor na direção A (bytes por segundo).

criação de várias sessões UDP; por outro lado, a região D abrange eventos com variação de energia menor, logo neste caso foram criadas menos sessões UDP. No segmento de altas frequências, as regiões E, F e G apresentam semelhanças com as regiões D, E e F da Figura 5.103, respetivamente. Os fluxos 4,7,8 e 16 apresentam uma variação de energia bastante alta, logo a taxa de receção de pacotes por parte de cliente é elevada. Isto alude que o cliente esteja a jogar jogos que requeiram tempo de resposta

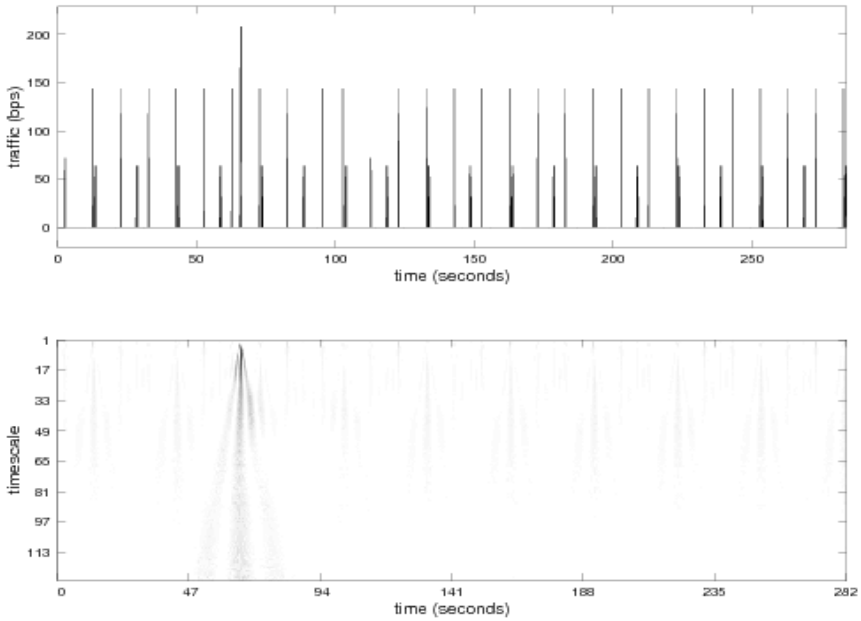

Figura 5.106 - Tráfego *upstream* XBOX por parte do servidor na direção B (bytes por segundo).

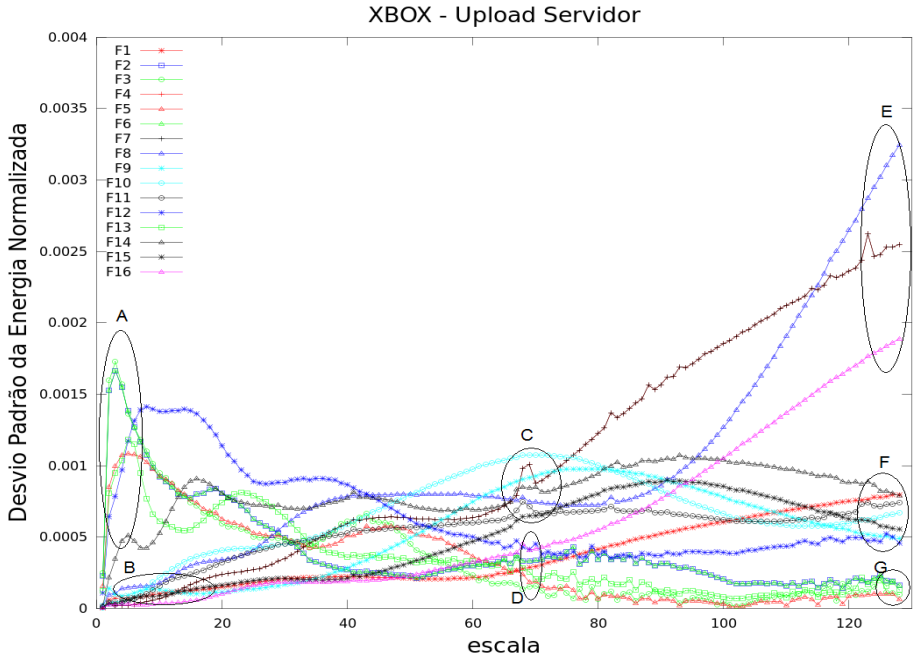

Figura 5.107 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* XBOX (do ponto de vista do servidor).

extremamente rápido durante um espaço de tempo relativamente alargado, como são os casos dos jogos de desporto e dos jogos de ação em primeira pessoa.

5.7.3 Cliente (*Upstream*)

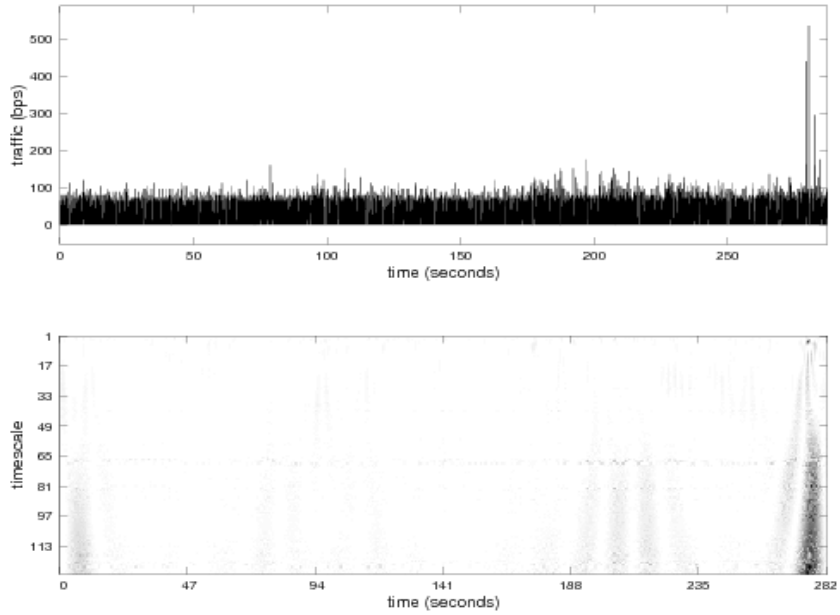

Figura 5.108–Tráfego *upstream* XBOX por parte do cliente na direção B (bytes por segundo).

Analisando a Figura 5.108, verifica-se que o volume de tráfego *upstream* é contínuo, o que se compreende tendo em conta as especificidades das plataformas de jogos online, em que necessário controlo em tempo real da ligação de cada jogador, de modo a não haver falhas. No final da amostra surgem picos de tráfego de curta duração e baixa amplitude, que correspondem a comandos do utilizador.

A Figura 5.109 apresenta características diferentes das observadas na Figura 5.108, pois neste caso o tráfego *upstream* é pseudo periódico, com picos de curta duração e baixa amplitude. Este tráfego está associado a pequenas componentes de baixa e média frequência no escalograma. Normalmente, estes componentes estão associados a pedidos do utilizador e ao estabelecimento de sessões, portanto é possível que neste caso o utilizador esteja a interagir com os vários menus da plataforma.

O estudo da Figura 5.110 permite concluir que as regiões A, B, C, D, F e G têm aspetos semelhantes às regiões homónimas da Figura 5.107, sendo que neste caso o tráfego *upstream* provém da aplicação do lado do cliente contrariamente ao cenário da Figura 5.107 em que o tráfego é originário do lado do servidor. A região E abrange dois fluxos (fluxos 1 e 6) com grande variação de energia e taxa elevada de transmissão de pacotes por parte do cliente, o que indica que este esteja a efetuar muitos cliques, logo a jogar um jogo que requeira atividade intensa por parte do cliente. A região F, tendo uma taxa de transmissão de pacotes mais reduzida, poderá estar associada a jogos em que o utilizador não seja tão ativo e onde haja vários momentos de pausa entre as suas ações (por exemplo jogos de questionários).

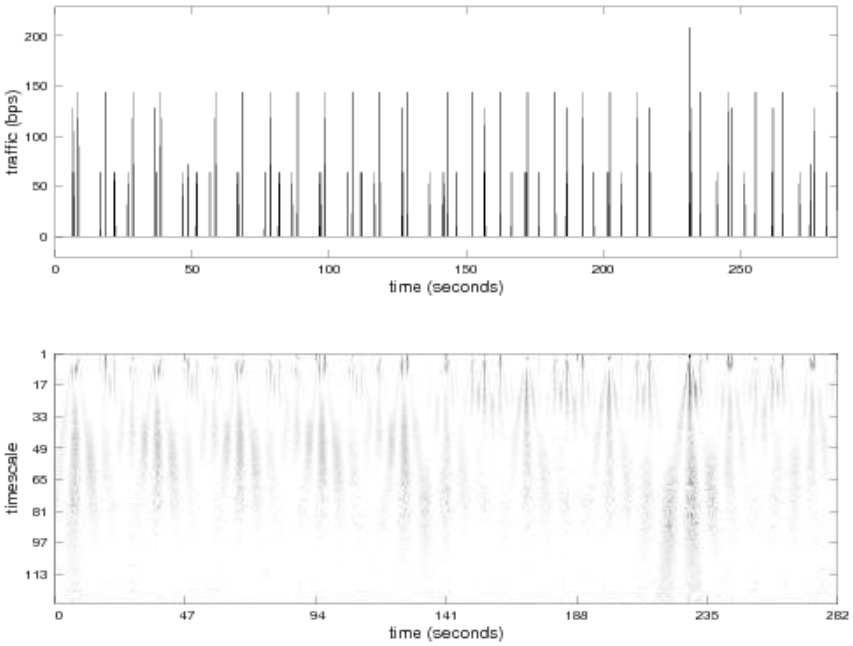

Figura 5.109 - Tráfego *upstream* XBOX por parte do cliente na direção B (bytes por segundo).

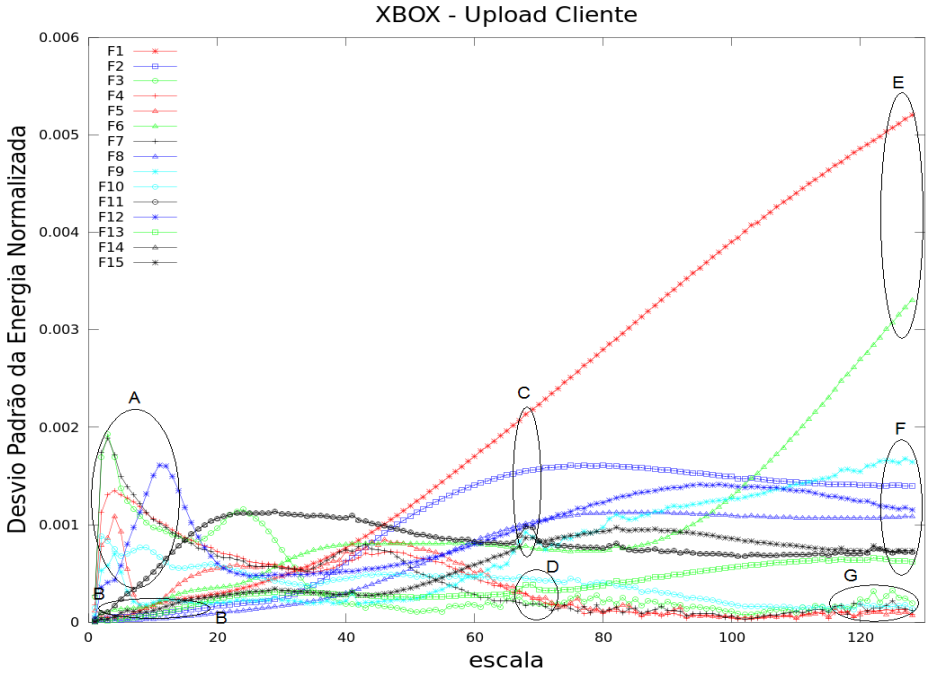

Figura 5.110 -Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* XBOX (do ponto de vista do cliente).

5.7.4 Servidor (*Downstream*)

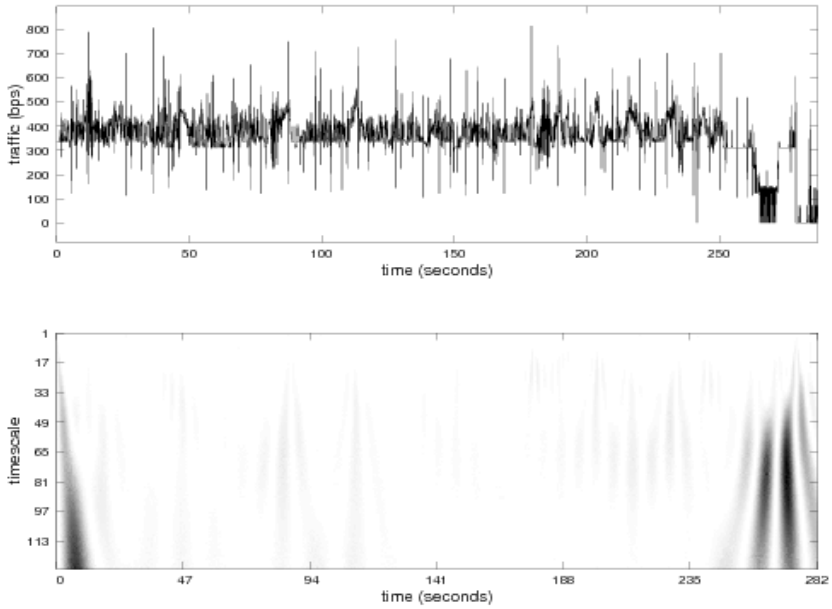

Figura 5.111 - Tráfego *downstream* XBOX por parte do servidor na direção B (bytes por segundo).

As seguintes quatro figuras representam casos diferentes de tráfego enviado pelos utilizadores, visto do porto de destino (por vezes este porto de destino pode ser o porto de serviço XBOX). A Figura 5.111 apresenta tráfego contínuo de carácter não periódico e instável de pacotes ao longo do intervalo temporal da amostra. A existência de componentes de média e alta frequência indicam que o utilizador está em atividade intensa.

Já a Figura 5.112 apresenta no seu primeiro minuto dois picos de tráfego muito pronunciados, com componentes de média e alta frequência de grande amplitude, o que alude a uma forte atividade por parte do utilizador. Depois, durante cerca de um minuto não há qualquer tráfego, que poderá dever-se a uma falha na ligação ou interrupção voluntária de uso do serviço voluntariamente. No fim deste minuto sem tráfego, este é retomado e apresenta um carácter não periódico e inconstante e apresenta componentes de alta frequência de baixa amplitude, o que indicia alguma atividade do utilizador.

No que diz respeito à Figura 5.113, o tráfego tem um carácter pseudo periódico, apesar de haver um intervalo de cerca de um minuto sem qualquer tráfego. Os picos de tráfego estão associados a pequenos componentes de baixa frequência, resultantes dos cliques de utilizador. Tendo em conta o tamanho dos pacotes em causa, o utilizador não estará a jogar, mas apenas a interagir com os menus da plataforma.

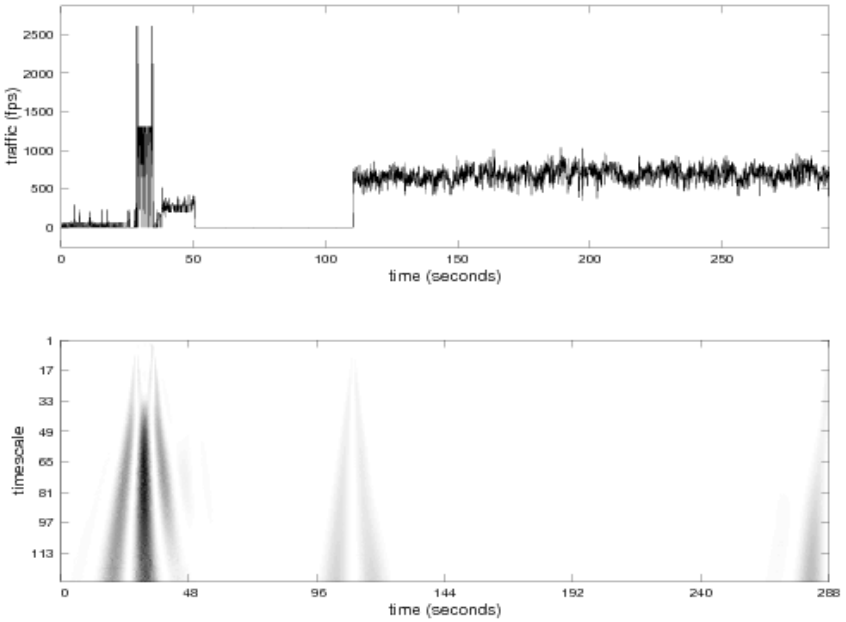

Figura 5.112 - Tráfego *downstream* XBOX por parte do servidor na direção A (bytes por segundo).

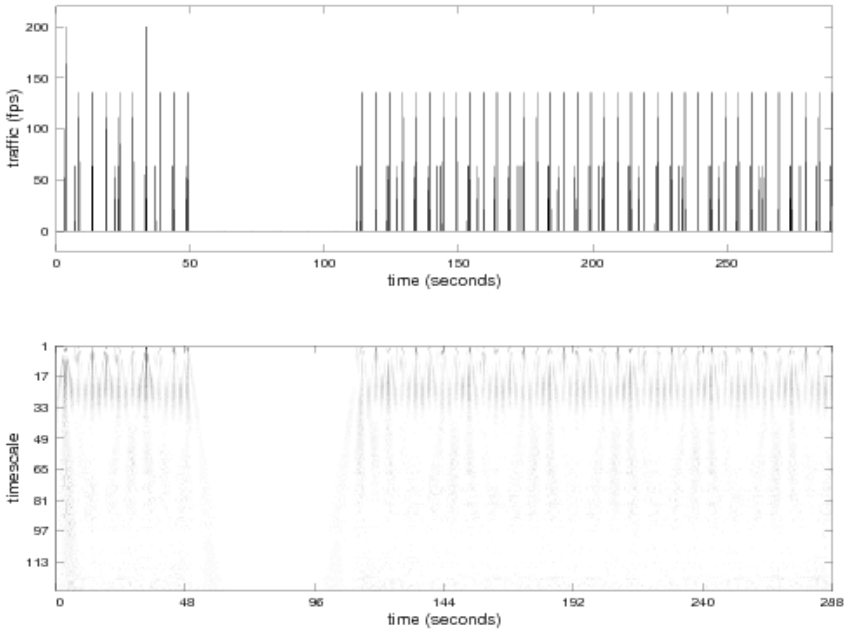

Figura 5.113 - Tráfego *downstream* XBOX por parte do servidor na direção A (bytes por segundo).

Finalmente, o tráfego apresentado na Figura 5.114 tem um carácter contínuo e constante, com muitos componentes de baixa frequência, o que indica que o utilizador está a jogar de forma intensiva e sem pausas.

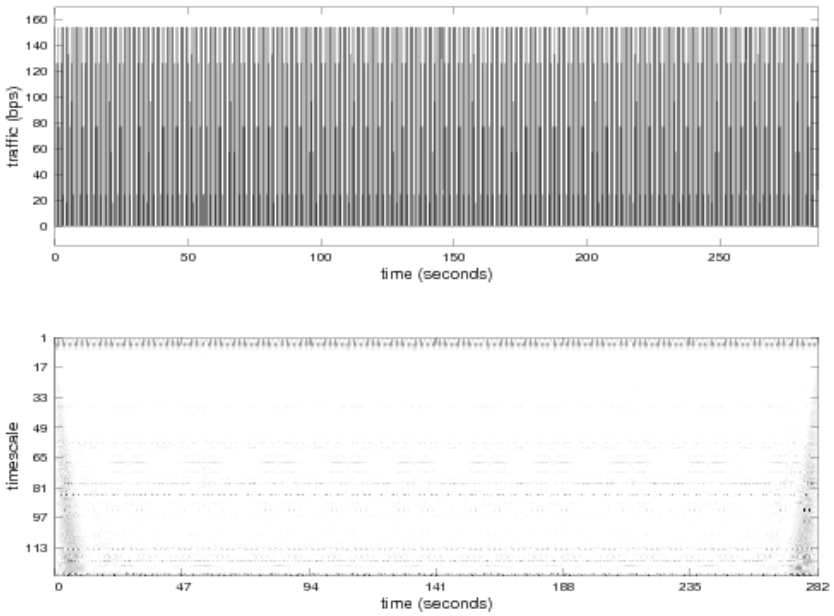

Figura 5.114 - Tráfego *downstream* XBOX por parte do servidor na direção B (bytes por segundo).

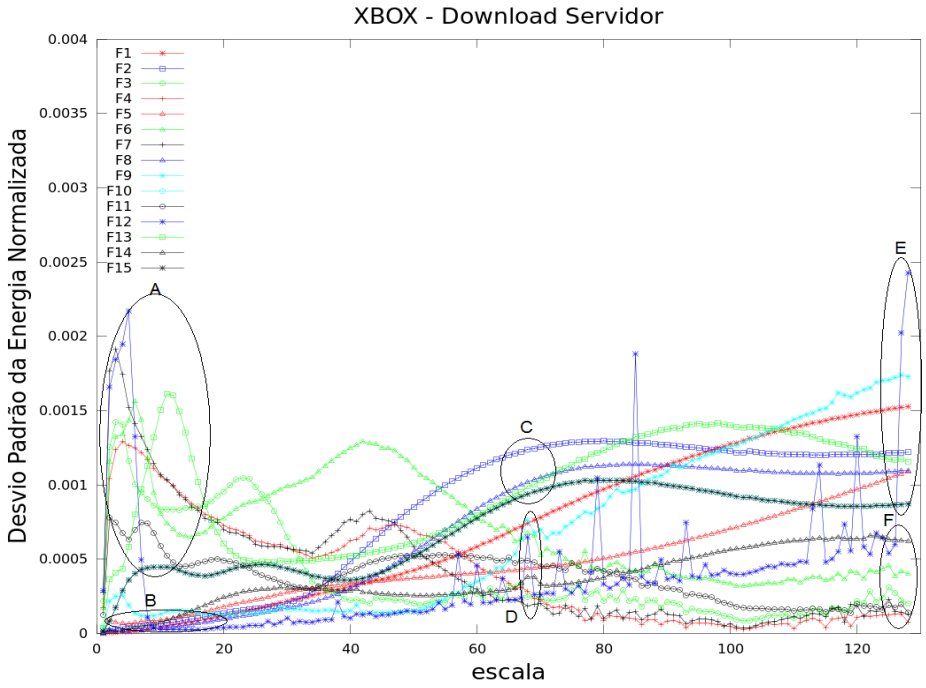

Figura 5.115 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* XBOX (do ponto de vista do servidor).

O estudo da Figura 5.115 permite concluir que as regiões A e B, no segmento de baixas frequências têm aspetos semelhantes às regiões homónimas da Figura 5.110, sendo que neste caso o tráfego é observado do ponto de vista do servidor. A região D abrange grande parte dos fluxos envolvidos neste cenário que apresentam variação de energia reduzida, dado que há criação de poucas sessões UDP nestes fluxos de tráfego. Por outro lado, a região C envolve fluxos de tráfego com variação de energia

considerável, pois nestes casos há mais interações UDP. No segmento de altas frequências, a região F compreende eventos com taxas de transmissão de pacotes reduzidas enquanto a região E abrange eventos com variação de energia de amplitude considerável, responsáveis por taxas de transmissão de pacotes elevadas, normalmente associados a jogos que requerem mais ações do jogador (como referido nas secções imediatamente anteriores). É importante realçar o comportamento do fluxo 12, pois demarca-se dos restantes fluxos pelo seu traçado irregular, com vários picos em diferentes segmentos da gama de frequências. Apresenta percentagens elevadas de componentes de baixas, médias e muito altas frequências, portanto está associado a tráfego com elevada taxa de transmissão de pacotes e que gera várias sessões de Internet ao longo do tempo.

5.8 Comparação Entre Protocolos

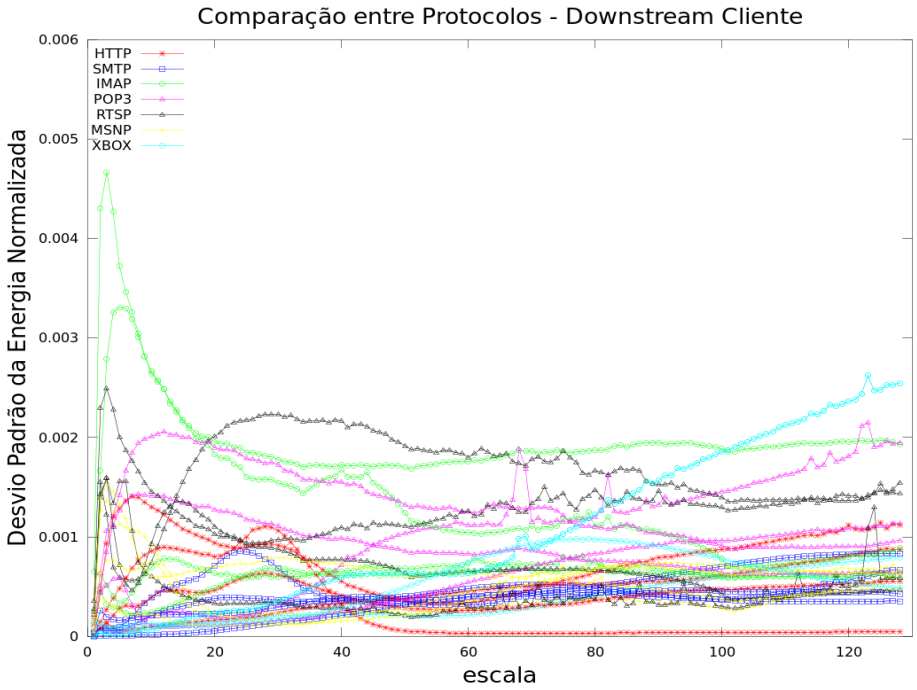

Figura 5.116 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *downstream* gerados por aplicações de diferentes protocolos (do ponto de vista do cliente).

Para efetuar-se a comparação do desvio padrão da energia dos fluxos de tráfego de diferentes protocolos, foram escolhidos quatro fluxos de tráfego de cada protocolo. Assim, analisando a Figura 5.116, verifica-se que são os fluxos gerados pelo protocolo IMAP que possuem uma maior variação de energia nas baixas frequências, seguidos pelos fluxos gerados pelos protocolos RTSP, MSNP, POP3 e HTTP. Os fluxos dos restantes protocolos contêm eventos com variação de energia diminuta nesta região. Ao nível das médias frequências, são de novo os fluxos destes dois protocolos que geram eventos com maior variação de energia. Os protocolos XBOX e POP3 também geram fluxos com variação de energia considerável neste segmento de frequência. Por outro lado, são os fluxos relacionados com o protocolo HTTP que apresentam menor variação de energia neste segmento de frequência. Ao nível das altas frequências, os protocolos RTSP e POP3 apresentam de forma mais consistente os fluxos com maior percentagem de componentes de alta frequência (apenas um fluxo do protocolo XBOX e outro do protocolo IMAP apresentam variação de energia de maior amplitude). Por outro lado, os protocolos SMTP e HTTP apresentam os fluxos com menor variação de energia neste segmento de frequência.

A Figura 5.117 analisa o tráfego *upstream* enviado pelo cliente para os diferentes protocolos. Verifica-se que os fluxos relacionados com os protocolos MSNP e RTSP apresentam maior variação de energia no segmento de baixas frequências; os fluxos gerados pelos protocolos XBOX, HTTP e SMTP mostram variação de energia considerável enquanto os restantes protocolos possuem fluxos com amplitudes de variação de energia diminutas. Ao nível do segmento de média frequência, os protocolos cujos fluxos possuem maior amplitude de variação de energia são XBOX, IMAP, POP3 e

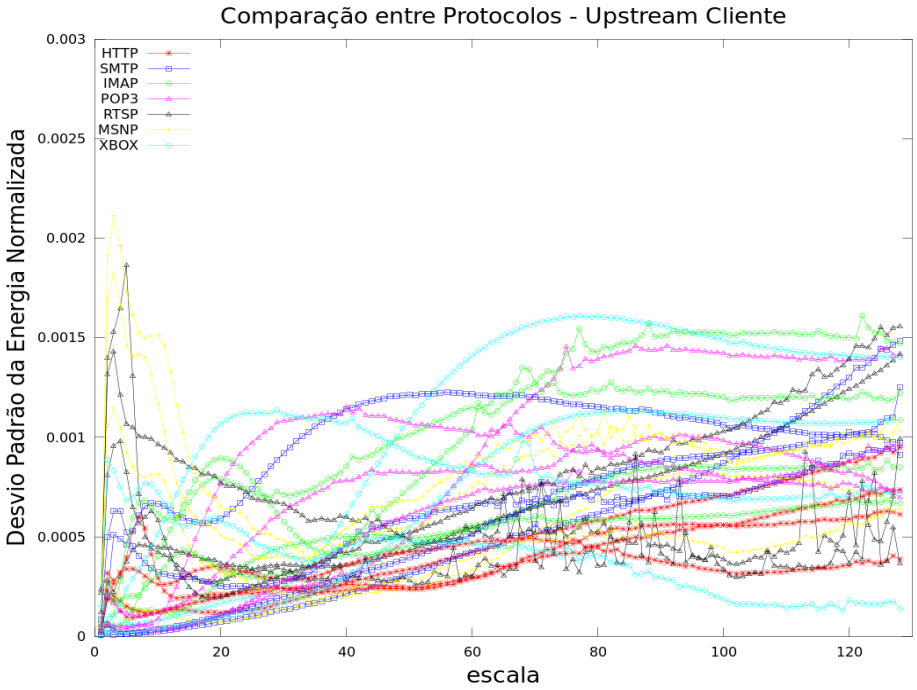

Figura 5.117 - Gráfico do desvio padrão da energia de vários fluxos de tráfego *upstream* gerados por aplicações de diferentes protocolos (do ponto de vista do cliente).

SMTP (apesar de ser apenas um fluxo, neste caso específico); no pólo oposto, os fluxos dos protocolos HTTP, XBOX, MSNP e SMTP (na sua maioria) geram eventos com variação de energia menor. Por fim, no segmento de altas frequências, são essencialmente os fluxos dos protocolos RTSP, IMAP, XBOX cujos eventos gerados contêm percentagens mais elevadas de componentes de alta frequência; por outro lado são sobretudo os fluxos dos protocolos RTSP, MSNP e HTTP cujos eventos apresentam variação de energia menor nesta zona.

6 Conclusões

O estabelecimento da Internet como um veículo preferencial de conteúdos em massa acarreta muitos desafios ao nível de uma gestão eficaz da disponibilização desses recursos ao dispor de milhões de pessoas a nível global. A Internet é utilizada como ferramenta de trabalho, consulta de notícias, lazer, para efetuar compras ou para transferências de ficheiros. Este crescimento no número de aplicações e serviços, bem como no número de utilizadores gerou uma concorrência cada vez mais feroz entre os ISPs, que se viram obrigados a aumentar a capacidade e qualidade dos serviços que oferecem. Esta concorrência, embora benéfica para os utilizadores, teve um grande impacto ao nível da gestão e performance das redes. Tornou-se então imperativo efetuar um mapeamento do tráfego da Internet associando-o a diferentes aplicações e serviços, para determinar-se que tipo de recursos e requisitos estão atribuídos a cada um. É assim explicado a necessidade da criação de perfis de tráfego e utilizador que sejam o mais rigorosos e completos possível.

Foi explorado o facto das aplicações interagirem de formas distintas com os utilizadores. Os pedidos dos utilizadores iniciam sessões protocolares, em que cada uma destas sessões gera séries de pacotes. Ao longo destas sessões criadas a partir dos pedidos dos utilizadores ocorrem eventos com diferentes componentes de frequência, que se distribuem da seguinte forma ao longo do espetro de frequências: baixas frequências (criação das sessões), frequências médias (tráfego ao longo da sessão) e frequências altas (chegada de pacotes). A diferenciação multi-escalar do tráfego capturado permite descrever as suas diferentes componentes de frequência, construindo assim uma “impressão digital” única de cada aplicação, que a distingue de todas as outras. Para a construção desta “impressão digital” também contribuem os escalogramas, que descrevem a energia do sinal ao longo do tempo e que são obtidos a partir da decomposição multi-escalar. Este método tornou-se adequado para ser utilizado neste trabalho, tendo em conta os constrangimentos e condicionantes das capturas de tráfego utilizadas no âmbito desta dissertação, nomeadamente ao nível da anonamização do tráfego; assim, este método permite contornar entraves relacionados com a privacidade e a encriptação dos dados em estudo.

Através da análise dos fluxos de tráfego gerados pelos diferentes protocolos estudados foi possível extrair algumas conclusões. No caso do protocolo HTTP, verificou-se que vários fluxos de tráfego apresentam características de perfil bastante similares, mesmo sendo gerados por aplicações diferentes. Aquando da comparação do tráfego anonamizado com o tráfego gerado por cinco classes de serviço distintas, foi possível perceber que alguns fluxos em estudo tinham características em comum com os fluxos das diferentes classes de serviço. Contudo, vários fluxos apresentaram características muito específicas, que não lhes permitem serem associados a nenhuma dessas classes de serviço e como tal apenas poderiam ser classificados recorrendo a algoritmos de classificação.

No que diz respeito aos protocolos utilizados nas aplicações de email, no caso do SMTP observou-se que o tráfego *upstream* enviado pelo cliente é mais volumoso que o tráfego *downstream* com destino ao cliente, o que se explica pelo facto deste ser um protocolo preferencial usado por aplicações de email para o envio de emails por parte do cliente e pela sua própria morfologia em que o email circula por diferentes servidores até ser entregue ao seu destinatário. Comparando o tráfego *downstream* destinado ao cliente dos protocolos IMAP e POP3, verificou-se que no caso do IMAP os fluxos com uma taxa elevada de transmissão de pacotes distinguem-se facilmente dos fluxos em que são transmitidos poucos pacotes, enquanto no caso do POP3 essa diferença não é tão

notória. Isto pode dever-se ao facto do IMAP promover um acesso remoto às operações nas caixas de email mais completo, relativamente ao POP3. Já o tráfego *upstream* proveniente do cliente tem um grafismo semelhante nos dois protocolos, em que a gama de taxas de transmissão de pacotes em ambos os casos é bastante grande.

Para o caso do protocolo RTSP, tornou-se complicado encontrar um comportamento padrão para os fluxos na situação do tráfego *downstream* destinado ao cliente, devido à disparidade dos perfis de cada fluxo. Isto deve-se ao facto deste protocolo ser usado por diferentes aplicações para transmissão de *streams* além de outros fatores como a qualidade do vídeo, velocidade da ligação à Internet ou a quantidade de utilizadores (*peers*) a visualizarem o mesmo conteúdo (a flutuação dos *peers* conectados dos quais se recebe dados e as suas conexões/desconexões influenciam a variação na qualidade do conteúdo recebido).

Para o caso do protocolo MSNP, foi possível encontrar um comportamento padrão dos fluxos de tráfego, tanto *downstream* como *upstream,* pois este protocolo gera tráfego muito específico: muitos cliques por parte do cliente, poucas sessões abertas e taxas de transmissão de pacotes normalmente moderadas.

Finalmente, no que concerne ao protocolo XBOX verificou-se que tanto no tráfego *downstream* como no tráfego *upstream* os fluxos apresentam perfis distintos, devido à diversidade de jogos que este protocolo suporta e que portanto requerem respostas diferentes por parte do cliente.

A Tabela 6.1 e a Tabela 6.2 apresentam os requisitos de QdS mais relevantes para cada aplicação estudada neste trabalho e que os ISPs e gestores de rede devem ter em conta no melhoramento das condições dos serviços que oferecem, respetivamente. As aplicações relacionadas com a interação em redes sociais, consulta de emails, partilha de fotos online e *instant messaging* não têm requisitos de tempo real e como tal o atraso e o tempo de espera associados a estas aplicações não são problemáticos. A questão de QdS mais relevante nestes casos prende-se com o número de clientes que utilizam estas aplicações durante determinados períodos de tempo. Tendo em conta o número crescente de utilizadores das redes sociais, por exemplo, torna-se crucial que as aplicações relacionadas com as redes sociais estejam preparadas para suportar tráfego muito intenso. Assim, conclui-se que nas situações do *browsing* e consulta de email não há necessidade de gerir requisitos em tempo real como são o caso do atraso ou do tempo de espera, por exemplo.

A visualização de vídeos online implica que o tempo de carregamento dos vídeos seja o menor possível (normalmente quanto maiores forem os buffers, menor será o tempo de carregamento total do vídeo), assim como o *jitter* (variação do atraso). É importante realçar que quanto maior for o tempo de duração do vídeo, maior será o tempo de carregamento. Portanto, é indispensável manter alocação de largura de banda constante durante o tempo de carregamento dos vídeos, de modo a baixar o atraso e o tempo de espera. Tendo em conta a possibilidade de visualização de vídeos em alta definição que certos sites oferecem, é importante garantir que vários requisitos de QdS multimédia têm boa qualidade: cor, resolução de ecrã, taxa de compressão e qualidade de som. Estes requisitos também se aplicam às aplicações de *streaming,* com a agravante destas aplicações normalmente gerarem muito tráfego que tem de ser processado e transmitido em tempo real e como tal a sensibilidade a atrasos e interferências durante a transferência é grande. Nestas duas situações, é essencial que ao nível dos recursos da rede o tempo de resposta seja o mais rápido possível.

No que diz respeito à aplicação XBOX, sendo uma aplicação em que o cliente envia e recebe dados em tempo real é muito sensível ao atraso e a interferências na transmissão de dados. Assim, o tempo de espera e de resposta têm de ser o mais pequeno possível. Esta aplicação envolve um volume de tráfego bastante grande, portanto requer uma pré-alocação de largura de banda constante para que a comunicação entre o cliente, os servidores da aplicação e outros utilizadores não seja

Aplicações	Requisitos de QdS - Serviço
Redes Sociais	Tempo de espera médio
Partilha de Fotos	Tempo de espera médio
Visualização de Vídeo	Pouco atraso, Tempo de espera pequeno, *Jitter* (variação no atraso) pequeno, Largura de banda constante, Taxa de compressão alta, Resolução de ecrã, Cor, Qualidade de som
Consulta de Email	Tempo de espera médio
Aplicações de *Instant Messaging* (MSNP)	Tempo de espera médio
Streaming (RTSP)	Pouco atraso, Tempo de espera pequeno, *Jitter* (variação no atraso) pequeno, Largura de banda constante, Taxa de compressão alta, Resolução de ecrã, Cor, Qualidade de som
Jogos (XBOX)	Pouco Atraso, Tempo de espera pequeno, *Jitter* (variação no atraso) pequeno, Largura de banda constante, Prioridade alta, Resolução de ecrã, Cor, Qualidade de som

Tabela 6.1 – Aplicações estudadas neste trabalho e respetivos requisitos de QdS do ponto de vista do serviço.

Aplicações	Requisitos deQdS – Rede
Redes Sociais	Atraso médio
Partilha de Fotos	Atraso médio
Visualização de Vídeo	Pouco Atraso, Baixo Tempo de resposta médio, *Jitter* (variação no atraso) pequeno, Largura de banda elevada
Consulta de Email	Atraso médio
Aplicações de *Instant Messaging* (MSNP)	Atraso médio
Streaming (RTSP)	Pouco Atraso, Baixo Tempo de resposta médio, *Jitter* (variação no atraso) pequeno, Largura de banda elevada
Jogos (XBOX)	Pouco Atraso, Tempo de espera pequeno, Baixo Tempo de resposta médio, *Jitter* (variação no atraso) pequeno, Largura de banda constante, Prioridade alta

Tabela 6.2 - Aplicações estudadas neste trabalho e respetivos requisitos de QdS do ponto de vista da rede.

afetada. Tendo em conta a grande variedade de jogos que esta aplicação oferece, a largura de banda a alocar irá variar tendo em conta o tipo de jogo que o utilizador selecione, assim como as definições que o mesmo escolha (resolução de ecrã, velocidade do jogo, etc). É importante realçar também que esta aplicação deve ter uma prioridade extremamente alta, tendo em conta que a atividade do utilizador é resposta aos comandos emitidos por outros utilizadores ligados entre si.

Como trabalho futuro, existem vários aspetos abordados no âmbito desta dissertação que carecem de mais investigação. É necessário programar scripts que, recorrendo ao modelo proposto na Figura 1.1, permitam que os fluxos de tráfego sejam automaticamente classificados e lhes sejam atribuídos os parâmetros de QdS associados a cada aplicação.

Para melhorar os parâmetros de QdS atribuídos a cada aplicação, é importante obter o máximo de informação sobre as características do tráfego gerado por cada aplicação. Como tal, seria importante recolher capturas de tráfego em diferentes zonas geográficas, efetuar medições em diferentes fusos horários e também em diferentes alturas do dia para tentar perceber as mudanças no comportamento dos clientes conforme a variação destes parâmetros. Outra investigação relevante prende-se com o estudo do tráfego relacionado com aplicações da Internet que tiveram um crescimento muito grande nos últimos anos como o armazenamento de ficheiros na *cloud* (utilizando aplicações como a Dropbox e Skydrive), plataformas de jogos online (como foi abordado neste trabalho o caso do protocolo XBOX) e ainda programas de reprodução de vídeo.

Outro assunto de abordagem pertinente seria a captura de tráfego durante acontecimentos de grande exposição mediática, como eleições, eventos desportivos, festivais de música ou catástrofes naturais. Durante estes eventos, sites de notícias online e aplicações de redes sociais ou de *microblogging* experimentam súbitos picos de tráfego que podem durar entre horas e alguns dias, o que acarreta dificuldades no acesso a estes sites. Como tal seria importante estudar como se comporta o tráfego nestes acontecimentos específicos, de modo a perceber os recursos que estas aplicações necessitam assim como os parâmetros de QdS a otimizar; assim, seria possível melhorar também a capacidade da resposta da rede (do ponto de vista dos ISPs e gestores de rede) a grandes variações no volume de tráfego num intervalo de tempo relativamente pequeno.

Portanto, o modelo de classificação de tráfego utilizado no âmbito desta dissertação permitiu diferenciar vários fluxos de tráfego gerados por diferentes protocolos, agrupando-os conforme o seu comportamento. Assim, este modelo conjugado com a utilização de algoritmos de classificação permite criar perfis de tráfego e de utilizador eficazes, bem como identificar os recursos requeridos por cada classe. Assim, será possível o desenvolvimento de parâmetros QdS mais adequados a cada aplicação.

7 Referências Bibliográficas

[1] M. M. Group. (2012, 17-08-2012). *Internet Growth Statistics*. Available: http://www.internetworldstats.com/emarketing.htm

[2] C. VNI. (2012, 17-08-2012). *Cisco Visual Networking Index: Forecast and Methodology, 2011-2016*. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[3] E. Rocha, P. Salvador, and A. Nogueira. Internet Users Multi-Scale Profiling for Network Management Purposes.

[4] K. Ramantas and K. Vlachos, "A TCP-Specific Traffic Profiling and Prediction Scheme for Performance Optimization in OBS Networks," *Journal of Optical Communications and Networking*, vol. 3, pp. 924-936, 17-11-2011 2011.

[5] N. M. M. Garcia, Paulo M. and M. M. Freire, "Measuring and Profiling IP Traffic," in *4th European Conference on Universal Multiservice, ECUMN'07*, Toulouse, France, 2007, pp. 283-291.

[6] K. C. Claffy, H. W. Braun, and G. C. Polyzos, "A parameterizable methodology for Internet traffic flow profiling," *Selected areas in Communications, IEEE Journal on*, vol. 13, pp. 1481-1494, 1995.

[7] K. Xu, F. Wang, S. Bhattacharyya, and Z.-L. Zhang, "A Real-time Network Traffic Profiling System," in *37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2007*, Edinburgh, United Kingdom, 2007, pp. 595-605.

[8] J. A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, "Creating evolving user behavior profiles automatically," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 24, pp. 854-867, 2012.

[9] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Googling the internet: Profiling internet endpoints via the world wide web," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, pp. 666-679, 2010.

[10] (2012, 08-21012). *RIAA*. Available: http://www.riaa.com/

[11] (2012, 10-2012). *Motion Picture Association of America*. Available: http://www.mpaa.org/

[12] H. Kim, K. C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Y. Lee, "Internet Traffic Classification Demystified: Myths, Caveats and the Best Practices," in *ACM SiGCOMM CoNEXT 2008*, Madrid, Spain, 2008.

[13] (2012, 20-2012). *OSI Model Reference Guide*. Available: http://compnetworking.about.com/cs/designosimodel/a/osimodel.htm

[14] E. Rocha, P. Salvador, and A. Nogueira, "Can Multiscale Traffic Analysis be Used to Differentiate Internet Applications?," *Telecommunications Systems*, vol. 48, pp. 19-30, September 2011 2012.

[15] (2012, 10-2012). *IANA - Internet Assigned Numbers Authority*. Available: http://www.iana.org/

[16] A. Dainotti, A. Pescapè, and K. C. Claffy, "Issues and future directions in traffic classification," *Network, IEEE*, vol. 26, pp. 35-40, 2012.

[17] H. Dreger, A. Feldmann, M. Mai, V. Paxson, and R. Sommer, "Dynamic application-layer protocol analysis for network intrusion detection," 2006, pp. 257-272.

[18] A. Moore and K. Papagiannaki, "Toward the accurate identification of network applications," *Passive and Active Network Measurement*, pp. 41-54, 2005.

[19] A. Madhukar and C. Williamson, "A longitudinal study of P2P traffic classification," 2006, pp. 179-188.

[20] (2012, 10-2012). *The Pirate Bay copyright crackdown is unsustainable*. Available: http://www.guardian.co.uk/commentisfree/2012/may/01/pirate-bay-copyright-crackdown

[21] A. Callado, C. Kamienski, G. Szabó, B. Gero, J. Kelner, S. Fernandes, and D. Sadok, "A survey on internet traffic identification," *Communications Surveys & Tutorials, IEEE*, vol. 11, pp. 37-52, 2009.

[22] P. Haffner, S. Sen, O. Spatscheck, and D. Wang, "ACAS: automated construction of application signatures," 2005, pp. 197-202.

[23] S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of p2p traffic using application signatures," 2004, pp. 512-521.

[24] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: multilevel traffic classification in the dark," 2005, pp. 229-240.

[25] (10-2012). *IPsec Tunneling*. Available: http://technet.microsoft.com/en-us/library/cc811544%28v=ws.10%29.aspx
[26] H. Kim, K. C. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Y. Lee, "Internet traffic classification demystified: myths, caveats, and the best practices," 2008, p. 11.
[27] (03-2012). *CoralReef Software Suite*. Available: http://www.caida.org/tools/measurement/coralreef/
[28] J. Erman, M. Arlitt, and A. Mahanti, "Traffic classification using clustering algorithms," 2006, pp. 281-286.
[29] Y. Chung, M. H. Park, and E. H. Paik, "A QoS negotiable service framework for multimedia services connected through subscriber networks," in *Consumer Electronics, 2006. ISCE'06. 2006 IEEE Tenth International Symposium on*, 2006, pp. 1-4.
[30] A. N. Mani, Arun. (2002, 12-09-2012). *Understanding quality of service for Web services*. Available: http://www.ibm.com/developerworks/webservices/library/ws-quality/index.html
[31] T. Kusano, T. Saydam, and S. Yucel, "Mapping QoS parameters for multimedia services using ATM MIB objects," in *Communications, 1998. ICC 98. Conference Record. 1998 IEEE International Conference on*, 1998, pp. 1425-1430.
[32] (2012, 03-10-2012). *A Short Overview of QoS Mechanisms and Their Interoperation*. Available: http://technet.microsoft.com/en-us/library/bb742477.aspx
[33] Webopedia. (2012, 11-2012). *MPLS*. Available: http://www.webopedia.com/TERM/M/MPLS.html
[34] Metaswitch. (11-2012). *What is MPLS and GMPLS?* Available: http://network-technologies.metaswitch.com/mpls/what-is-mpls-and-gmpls.aspx
[35] L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource reservation protocol (RSVP)--Version 1 functional specification," *Resource*, 1997.
[36] C. Metz. (1999, May-June 1999) RSVP: General-Purpose Signaling for IP. *IEEE Internet Computing*. 95-99.
[37] P. P. White. (1997, May 2007) RSVP and Integrated Services in the Internet: A Tutorial. *IEEE Communications Magazine*. 100-106.
[38] Microsoft. (2012, 04-10-2012). *A Short Overview of QoS Mechanisms and Their Interoperation*. Available: http://technet.microsoft.com/en-us/library/bb742477.aspx
[39] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, "An architecture for differentiated services," 1998.
[40] P. A. Morettin, "From Fourier to wavelet analysis of time series," *Proceedings in Computational Statistics*, pp. 111-122, 1996.
[41] P. Salvador and A. Nogueira. Differentiating Users Hidden Behaviors with Traffic Scalogram Analysis.
[42] W. Willinger, "The discovery of self-similar traffic," *Performance Evaluation: Origins and Directions*, pp. 513-527, 2000.
[43] I. Daubechies, *Ten lectures on wavelets* vol. 61: SIAM, 1992.
[44] E. O. E. Rocha, "Methodologies for traffic profiling in communication networks," PhD, Universidade de Aveiro, 2011.
[45] M. Fomenkov and K. C. Claffy, "Internet Measurement Data Management Challenges," in *Workshop on Research Data Lifecycle Management*, Princeton, New Jersey, USA, 2011.
[46] CAIDA. (2012, 03-2012). *CAIDA - The Cooperative Association for Internet Data Analysis*. Available: http://www.caida.org/home/
[47] CAIDA. (2012, 03-2012). *Passive Monitor: equinix-chicago*. Available: http://www.caida.org/data/monitors/passive-equinix-chicago.xml
[48] (2012, 10-2012). *What Time Is It Around The World Right Now?* Available: www.24timezones.com
[49] J. Klensin. (2008). *Simple Mail Transfer Protocol - RFC 5321*. Available: http://datatracker.ietf.org/doc/rfc5321/
[50] Microsoft. (2003, 10-2012). *How POP3 Service Works*. Available: http://technet.microsoft.com/en-us/library/cc737236%28v=ws.10%29.aspx
[51] M. C. Brain, Tim. (2007, 10-2012). *How E-mail Works*. Available: http://computer.howstuffworks.com/e-mail-messaging/email.htm
[52] M. G. Syme, Philip. (2004, 10-2012). *Understanding Application Layer Protocols*. Available: http://www.informit.com/articles/article.aspx?p=169578&seqNum=3
[53] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (RTSP) RFC 2326," *IETF (April 1998)*, 1998.
[54] S. Cooper. (10-2012). *MSNP Protocol*. Available: http://www.ehow.com/facts_7510149_msnp-protocol.html

[55] (2003, 10-2012). *MSN Messenger Protocol*. Available: http://www.hypothetic.org/docs/msn/general/overview.php
[56] Microsoft. (2012, 10-2012). *XBOX LIVE - Xbox.com*. Available: http://www.xbox.com/pt-BR/Live
[57] I. S. Petiz, "Caracterização de tráfego e comportamentos numa rede P2P-TV," MSc, Universidade de Aveiro, 2010.